NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

Canada

X CHROMOSOME INACTIVATION AND THE PHOSPHOGLYCERATE KINASE GENE FAMILY

by

Chaker Nadim Adra

A thesis presented to the University of Ottawa in partial fulfillment of the requirements for the degree of Doctor of Philosophy

ISBN 0-315-46705-3

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

PREFACE

I believe that there is a need to replace competition by cooperation in our scientific studies and to introduce immediate reforms. Graduate students and postdocs are not well paid and are often required to work very long hours. Research is similar to the olympic games. As an example the work achieved and described in this thesis represents goals towards which many laboratories around the world have strived. I am very sorry that our group was the first to achieve these goals thus forcing the other groups to stop their programs after investing several years of research.

DEDICATION

I dedicate my work to my mother Sheikha Wahida Adra, my father Sheikh Nadim Jawad Adra, to my sisters Sheikhas Najwa and May, and to my brothers Sheikhs Jawad, Ghassan, Mahmoud and Hicham. My parents dedicated their lives to the family, to people and to the issue of peace. They have always worked to make Lebanon a better place and to improve relations between communities.

Also, I dedicate this work to my new Country, Canada, which accepted me when I arrived at Mirabel Airport without a visa, in 1976, from a war torn country. During those difficult days, I didn't know that I was destined to become a scientist.

This thesis describes discoveries that began over 6 years ago, discoveries that involved the support of family and friends and the usage of animals. I am very grateful to everyone. To the animals that I have sacrificed I always repeated in my heart the words of Gibran Khalil Gibran:

But since you must kill to eat, and rob the newly born of its mother's milk to quench your thirst, let it then be an act of workship,

And let your board stand an altar on which the pure and the innocent of forest and plain are sacrificed for that which is purer and still more innocent in man.

When you kill a beast say to him in your heart,

"By the same power that slays you, I too am slain; and I too shall be consumed.

For the law that delivered you into my hand shall deliver me into a mightier hand.

Your blood and my blood is naught but the sap that feeds the tree of heaven."

The University of Ottawa requires the signatures of all persons using or photocopying this thesis. Please sign below, and give the address and date.

ACKNOWLEDGMENTS

I would like to offer my thanks to all those who have helped me in many different ways.

To Dr. M.W. McBurney for his guidance.

To the members of my research committee, Drs. L.P. Visentin, P. Moreau, D. Hickey and H.C. Birnboim.

To the examiners of my Ph.D. thesis, Drs. J. Rossant, David Brown, Donn Kushner, H.C. Birnboim and V.N. Iyer.

To Jane Craig and Karen Jardine. Karen provided me with DNA from bacterial strains and from EC cell lines for the methylation studies. Her help was important in accelerating the rate of progress.

To the graduate students and post-doctoral fellows, J. Bell, Gary Paterno, Giovanna, Steve, Chantal, Martha, José, Mandy, Molly, Soma, René, Trevor, Bill, Randy, Poppo and Benoit. I will always remember Mike Rudnicki who was my classmate, colleague and a friend during all the undergraduate and graduate years. Mike has an open mind. He is very interested in current issues, often willing to listen and to give help.

To Jacque Helie and George Ben-Tchavtchavadze for their help with the figures.

To my family, the Natural Sciences and Engineering Research Council of Canada and the Medical Research Council of Canada for their financial support.

I thank Drs. A. Michelson and S. Orkin for their generous gifts of mouse and human pgk-1 cDNA clones. I am grateful to Drs. D.R. Cox, P. Ashley, F. Ruddle, D. Pravtcheva and P. D'Eustachio for

providing mouse-hamster somatic cells DNAs and to Drs. V. Chapman, D.R. Cox and J.R. McCarrey for helpful discussions.

I was very happy that I have met and collaborated with Drs. Y.-F. Lau and N. Ellis.

In 1985 my family provided me with money to participate at a satellite conference of the 10th International Congress of the Society of Developmental Biologists at Marina Del Ray. I wanted to participate in the meeting to establish contacts with scientists working in the field of spermatogenesis. There I met Dr. N.B. Hecht. Dr. Hecht, with a glass of wine in his hand, told me about his studies concerning the expression of pgK-2 during spermatogenesis. I told him that I am cloning the pgK-1 and the pgK-2 genes. He suggested that I should call Dr. Y.-F. Lau. At the end of the meeting I flew from Los Angeles to San Francisco. Dr. Jeanette Garretty, who was waiting for me at the airport, drove me to the Howard Hughes Medical Institute at the University of California in San Francisco. Dr. Lau came to meet us at the secretariat. He welcomed both of us smiling, shook our hands in a very affectionate way and invited us for a coffee. In his office I showed him my southern blots of genomic DNA and we made an agreement: 1) he will provide me with the mouse pgK-2 cDNA, 2) I will continue my work on the cloning of the pgK-1 and pgK-2 genes making use of the pgK-1 and pgK-2 cDNA probes. After our meeting I went to see Dr. D.R. Cox who accepted to provide me with somatic cell hybrids DNAs for the mapping studies. My trip to California has had results that could hardly have been anticipated. It led to the cloning of the human and the mouse pgK-2 genes. The pgK-2 gene is for the moment the sole example of a fully processed functional retrogene. This discovery suggests that retroposition may have been used as a mechanism for generation of functional genes.

TABLE OF CONTENTS

PAGE

PAGE

PAGE

List of Figures

List of Tables

Chapters I, III, IV and VI

no tables

Abbreviations

alpha-gal - gene encoding alpha galactosidase

AFP - alpha-fetoprotein

bp - base Pair(s)

BRL - Bethesda Research Laboratories

cDNA - DNA complementary to mRNA

DEPC - diethyl pyrocarbonate

EC - embryonal carcinoma

EDTA - ethylene diamine tetraacetic acid

EtdBr - ethidium bromide

5-AC - 5-azacytidine

G6PD - glucose-6-phosphate dehydrogenase

g6pd - gene encoding G6PD

HAT - hypoxanthine-methotrexate-thymidine

HAT^+ - grown in HAT medium

HPRT - hypoxanthine phosphoribosyl transferase

hprt - gene encoding HPRT

Kb - kilobase(s) or 1000 bp

LB - Luria-Bertani

MIF - mouse interspersed fragment

MOPS - 3-[N-morpholino]propane-sulfonic acid

mRNA - messenger RNA

nt - nucleotide(s)

oct - gene encoding ornithine carbamoyl transferase

PBS - phosphate buffered saline

PGK - phosphoglycerate kinase

pgk-1 - gene encoding the somatic cell isoform of PGK

pgk-2 - gene encoding the testis-specific PGK

RA - retinoic acid

SDS - sodium dodecyl sulfate

TCA - trichloroacetic acid

TE - tris-EDTA [10 mM tris-HCl, 1 mM EDTA]

UPE - upstream promoter element

Xa - active X chromosome

Xce - X controlling element

Xi - inactive X chromosome

Xic - X inactivation center

ABSTRACT

Female mammalian somatic cells usually contain two X chromosomes whereas each male cell has only one X chromosome. The expression levels of X-linked genes are equal in male and female tissues due to dosage compensation which is responsible for the presence of one active X chromosome in each female somatic cell. Because of strong evidence for the role of DNA methylation in X inactivation (Monk, 1986) and because of the fact that X inactivation can be analyzed in culture using female embryonal carcinoma (EC) cells (Paterno and McBurney, 1985), we decided to examine the molecular mechanisms involved in X inactivation by using DNA demethylating agents to reactivate the inactive X chromosomes in our EC cell lines (Paterno et al., 1985) and by cloning and studying the regulation of the mouse phosphoglycerate kinase (pgk-1) gene.

The mammalian genome contains two genes encoding phosphoglycerate kinase; the pgk-1 gene is X-linked and is expressed in all cells while the pgk-2 gene is expressed exclusively in spermatocytes and spermatids. The cloning of the pgk-1 gene was very difficult because of the presence of related sequences. We cloned all these sequences (Adra et al., 1988). We have shown that the mouse genome contains no pseudogenes derived from pgk-2. On the other hand, the genomes of Balb/c and C3H/He strain mice contain six other regions with sequences homologous to those of pgk-1 cDNA. These pgk-related sequences are likely derived from the pgk-1 gene by retroposition because all are located on autosomal chromosomes and because none appear to be interrupted by introns. Two of the presumed pseudogenes contain sequences homologous to all regions of the pgk-1 cDNA while the other four genomic regions were truncated at the 5', 3', or both ends. One

of the truncated pseudogenes was sequenced. Its pgk-related sequence was not flanked by direct repeats, suggesting that loss of the 5' and/or 3' ends of this retrogene may have occurred following its integration into the genome. Our evidence suggests that pgk-1-derived retroposons arose initially more than 100 million years ago and have continued to arise until so recently that some are unique to different mouse strains.

In addition, we cloned the mouse pgk-2 gene and mapped it to chromosome 17 (Adra et al., 1988). Our results also suggested that the pgk-2 gene arose from the pgk-1 gene by retroposition between 100 and 200 million years ago and that retroposition has resulted in the production of at least one functional gene (Boer et al., 1987).

We also cloned the pgk-1 gene (Adra et al., 1987). The gene is contained within a 16-kb region of the X chromosome and is interrupted by at least ten introns. The promoter region of the pgk-1 gene is rich in G and C nucleotides and contains five copies of the hexadeoxynucleotide, GGGCGG, the potential binding site for the Sp1 transcription factor, a CCAAT sequence, but no TATA box. This promoter functions following DNA-mediated transfection into mammalian cells. The promoter of the mouse pgk-1 gene is homologous to the human pgk-1 promoter. A number of conserved motifs in the promoter may indicate a significant role for these sequences in expression of the pgk-1 gene.

To test for a causative relationship between gene inactivity and methylation (reviewed by Bird, 1986) we investigated the methylation pattern at HpaII and HhaI sites present in the 5' control region of the mouse pgk-1 gene in kidney, spleen and EC cell lines representing different stages of X chromosome differentiation. Our analysis of the

upstream region indicated that the CCGG sites assayed are unmethylated on both the active and inactive X chromosomes. These results suggest that mechanisms other than methylation must be involved in maintaining the X chromosome silent. To verify this inference I have initiated the functional characterization of the promoter region with the hope that it will allow us to address questions concerning not only the regulation of the expression of the pgk-1 gene, but also the mechanism of X chromosome differentiation.

Résumé

Les cellules femelles somatiques de mammifères contiennent deux chromosomes X tendis que chaque cellule male contient un seul chromosome X. Les niveaux d'expréssion des gènes localisés sur le chromosome X sont égaux dans les cellules males et femelles à cause d'un mécanisms de compensation de dosage qui est responsable de la présence d'un seul chromosome X actif dans chaque cellule femelle somatique. A cause de certains résultats suggérant que la méthylation de l'ADN est impliquée dans le phénomène d'inactivation du chromosome X (Monk, 1986) et du fait que ce phénomène peut être analysé en utilisant les cellules de tératoearcinomes (EC) (Paterno et McBurney, 1985), nous avons décidé d'examiner les mécanismes moléculaires impliqués dans l'inactivation du chromosome X en utilisant des agents démithylant l'ADN pour réactiver le chromosome X inactif dans nos lignées EC (Paterno, Adra et McBurney, 1985) et an clonant et étudiant la régulation du gêne phosphoglycerate kinase 1 (pgk-1) localisé sur le choromosome X de la souris.

Le génome de mammifères contient deux gènes codant pour la phosphoglycerate kinase; le pgk-1 est localisé sur le chromosome X et exprimé dans toutes les cellules tandis que le gêne pgk-2 est exprimé exclusivement dans les spermatocytes et spermatides. Le clonage du pgk-1 était très difficile à cause de la présence des séquences homologues. Nous avons clonés toutes ces séquences (Adre et al., 1988). Nous avons montré que le génome de la souris ne contient pas de pseudogènes dérivés du pgk-2. Les génomes des souches de souris Balb/c et C3H/He contiennent six autres locus possédant des séquences homologues au pgk-1 cADN. Ces séquences sont probablement dérivées du pgk-1 par rétroposition car elles sont toutes localisées sur des

chromosomes autosomaux et apparament aucune de ces séquences no possède d'introns. Deux de ces pseudogènes contiennent des séquences homologues à toutes les régions de pgk-1 cADN tandis que les autres quatre régions génomiques sont trançonnées à l'extrémité 5' 3', ou les deux. Un de ces pseudogènes trançonnés a été séquencé. Sa séquence homologue au pgk-1 n'était pas limité par les éléments répétitifs directes suggérant que la perte de l'extrémité(s) 5' et/ou 3' de ce retrogène a possiblement eu lieu après sont intégration dans le génome. Notre évidence suggère que les rétroposons du pgk-1 sont apparus il y a plus de 100 millions d'années et que la rétroposition a persisté jusqu'à récemment au point que certains retroposons sont uniques à différentes souches de souris.

Egalement, nous avons cloné le gène pgk-2 et déterminé qu'il est localisé sur le chromosome 17 chez la souris (Adra et al., 1988). nos résultats suggérent que le pgk-2 est dérivé du pgk-1 par rétroposition datant de 100 à 200 millions d'années et que la rétroposition a produit au moins un gène fonctionnel (Boer, Adra et al., 1987).

Nous avons aussi cloné le pgk-1 (Adra et al., 1987). Ce gène s'étends sur une région de 16-Kb du chromosome X et il est interrompu par au moins 10 introns. Le promoteur du pgk-1 est riche en nucléotide G et C et contient 5 copies de l'hexadeoxynucléotide, GGGCGG, qui est un site potentiel d'attachement pour le facteur de transcription Sp1, une séquence CCAAT, mais ne contient pas une boîte TATA. Le promoteur fonctionne après transfection dans les cellules de mammifères. Il est homologue à celui du pgk-1 humain. La présence de motifs conservés dans ce promoteur est probablement un indicateur d'un role significatif de ces séquences dan sl'expression du pgk-1.

Afin de vérifier si l'inactivité génomique et la méthylation de

l'ADN partagent une relation de cause à effet (Bird, 1986) nous avons étudié la mode de méthylation aux sites HpaI et HpaII présents dans la région 5' du gène pgk-1 dans des tissues du rein, du foie et des cellules EC représentants diverses étapes de l'inactivation du chromosome X. Notre analyse du région 5' a indiqué que les sites CCGG examinés ne sont pas méthylés sur les chromosomes X actifs et inactifs. Nos résultats suggérent qu'il y a des mécanismes autre que la méthylation qui maintiennent le chromosome X inactivé. Pour tester cette hypothèse j'ai initié la caractérisation du côté 5' avec l'espoir de répondre à des questions concernant non seulement la régulation de l'expression du pgk-1 mais aussi les mécanismes de différentiation du chromosome X.

CHAPTER I

INTRODUCTION

Female mammalian somatic cells usually contain two X chromosomes whereas each male cell has only one X chromosome and the male determining Y chromosome. The expression levels of X-linked genes are equal in male and female tissues due to dosage compensation which is responsible for the presence of only one active X chromosome in each female somatic cell. The inactivation event accompanies cell differentiation during early embryogenesis. The inactivated X chromosome becomes genetically inert and cytologically heterochromatic; once established, the parental inactive state is maintained and stably inherited in all the mitotic descendants of somatic cells.

1.1 X Chromosome activity in germ cells or in vivo reactivation:

The absence of a highly condensed X chromosome suggested that oocytes of the quiescent dictyate stage have two active X chromosomes (Ohno, 1963).

Confirmation of this inference was obtained from the analyses of human and mouse oocytes that were heterozygous for electrophoretic glucose 6 phosphate dehydrogenase (G6PD) variants. These studies showed the presence of homo- and heterodimers of G6PD suggesting that both X chromoxomes are active in each cell (Gartler et al., 1975; 1980). Also, gene dosage studies demonstrated that oocytes of XO mice have half as much activity of X chromosome gene products as do oocytes

of XX mice (Epstein, 1969; 1972; Kozak et al., 1974; Mangia et al., 1975; Andina, 1978; Gartler and Rivest, 1983).

The observation that primary oocytes do contain two active X chromosomes suggests two possible events. The first is that germ cells arise from a separate embryonic cell pool that does not inactivate. The second possibility is that reactivation of the inactive X occurs during germ-cell differentiation.

The existence of a single active X chromosome in primordial germ cells was inferred from cytogenetic studies. These studies have detected a heterochromatic X chromosome in germ cells during their migration from the epiblast to the gonadal ridges (Ohno, 1964) and during their differentiation into oogonia (Gartler et al., 1980). As is mentioned in the next section, all of the evidence suggests that X inactivation has also occurred in most other cells of the female embryo before this stage (7.5 days gestation - 11.5 days gestion). These studies suggested that germ cells undergo X chromosome reactivation.

Studies of allozyme expression in germ cells from heterozygous mouse foetuses (G6PD, HPRT, PGK-1) confirmed the presence of a single active X chromosome in oogonia and placed the time of X chromosome reactivation just before the onset of meiosis (Andina, 1978; Johnston, 1981; Kratzer and Chapman, 1981; McMahon et al., 1981; Monk and McLaren, 1981).

This conclusion is also supported by the observation of an increase in mouse HPRT activity in XX compared to XO germ cells as they entered meiosis (Monk and McLaren, 1981; Andina, 1978; Mangia et al., 1979).

McMahon et al. (1981) used allelic variants of PGK-1 to study X chromosome inactivation and its randomness. They found that pre-meiotic germ cells from a heterozygous female mouse exhibited both variants in equal amounts. Thus, X chromosome inactivation is a random process in primitive germ cells as it is in somatic cells (see section 1.5).

Overall the inactivation-reactivation process is supported by cytogenetic, allozyme expression and gene dosage studies. Also these studies place the time of X chromosome reactivation just before the onset of meiosis.

1.2 X chromosome activity during preimplantation development:

A variety of procedures were used by different laboratories to determine whether or not dosage compensation exists in the female embryo before the occurrence of X inactivation. Cytogenetic studies did not reveal a heterochromatic or a late replicating X chromosome in early female embryos (Takagi, 1974) suggesting that both X chromosome are genetically active. In addition, four X-linked enzymes (HPRT, G6PD, PGK-1, alpha-gal) and related autosomal enzymes were used as controls to determine the frequency distribution of enzyme activities in single embryos at various developmental stages. A bimodal distribution was evident at the eight-cell stage and early blastocyst (2-5 days gestation). This distribution is thought to reflect the activities of X-linked genes in XY embryos and in female embryos with 2 active X chromosomes (Kratzer and Gartler, 1978; Monk and Harper, 1978). Thus, prior to X inactivation during these early

developmental stages, there is no dosage compensation.

1.3 X chromosome activity during embryogenesis:

A unimodal distribution considered to reflect the activities of XY embryos and of female embryos with one X chromosome active was first detectable in the mouse in the blastocyst stage onwards. Biochemical analysis of differences in specific activity of X-encoded enzymes have shown that X inactivation is not synchronous in all cells of the embryo (reviewed by Monk, 1978). At 3 days gestation, cells on the outer surface of the morula differentiate to form trophectoderm and an inactive X is first observed during that event (Kratzer and Gartler, 1978). The paternal X chromosome is preferentially inactivated (Takagi et al., 1974; 1975; 1978; also see section 1.5). At 4.5 days gestation, the formation of primitive endoderm is also accompanied by paternal X chromosome inactivation (Kratzer and Gartler, 1978). Implantation does not occur until after the formation of primary endoderm. The embryonic ectoderm of the inner cell mass continues to express both X chromosomes. Not until 6.5 days gestation is the differentiation of the embryonic ectoderm to form ectoderm, mesoderm and endoderm completed. This step is also accompanied by the inactivation of one X chromosome in each female cell (Monk and Harper, 1979; Takagi et al., 1982). These studies have shown that either the paternal or the maternal X chromosome can become inactivated in each cell of the embryonic ectoderm and that inactivation is random.

These observations indicated that X inactivation process appears to accompany an early step of cell differentiation (Monk and Harper, 1978).

1.4 Stability of X chromosome inactivation: A rule that has been violated:

Stable inactivation was demonstrated by investigating tumours in mammals heterozygous for electrophoretic variants of G6PD (Gartler and Andina, 1976), by assaying for HPRT activity in cells heterozygous for HPRT-deficiency and carrying the wild type hprt allele on the inactive X chromosome (Migeon, 1972; Romeo and Migeon, 1975; deJonge et al., 1982; Graves and Young, 1982) and by analysing cultured skin fibroblast cells from females heterozygous for several X-linked markers. Clones derived from single cells contained the parental G6PD homodimer bands (A or B) but not both (Davidson et al., 1963; Rattazi and Cohen, 1972; Ray et al., 1972). Since G6PD is a dimer, homo- and heterodimers should be evident if both variants are expressed in the same cell.

In summary, following X inactivation, the parental inactive state is maintained and stably inherited in all the mitotic descendants of somatic cells in vivo and in culture. It is important to note that very recently Wareham and co-workers (1987), using mice with Searle's X-autosome translocation demonstrated for the first time that the silent gene encoding ornithine carbamoyl transferase (OCT) on the inactive normal X chromosome is reactivated during aging.

1.5 Random and non-random X chromosome inactivation:

Lyon (1961) said in her hypothesis that either the paternal or the maternal X chromosome can become inactivated at random. However,

this postulate had to be modified later in light of the fact that inactivation in nonembryonic eutherian tissues (see section 1.2) and in methatherian mammals is non-random (for a review, see VandeBerg, 1983).

Preferential inactivation of the paternal X chromosome was observed in all the cells of extra-embryonic tissues. Takagi and his associates (1974; 1975; 1978) followed the fate of the paternal and the maternal X chromosomes morphologically in mouse conceptuses carrying the Cattanach's translocation T(X;7) 1 Ct which involves the insertion of a fragment of chromosome 7 into the X chromosome (Ohno and Cattanach, 1962), rendering the T(X;7) chromosome longer than the intact X. They found that the paternally derived X chromosome was always allocyclic in the yolk sac and the chorion whereas approximately equal numbers of embryonic cells contained late replicating paternal or maternal X chromosomes (Takagi and Sasaki, 1975). Takagi (1978) also noted that the inactive X chromosome in extraembryonic tissues of day 5.3 embryos replicated earlier than the active X chromosome. This pattern of early replication shifted between day 6 and 6.5 so that the early replicating inactive X chromosome becomes late replicating (Takagi et al., 1982).

More evidence for the non-random inactivation of the paternal X chromosome came from biochemical studies. These studies demonstrated that the allocyclic paternal X chromosome is functionally inactive by using allozyme variants of PGK-1. The maternally-derived PGK-1 was expressed in the trophectoderm and primitive endoderm lineages in mouse whereas embryonic lineages were mosaic for X-chromosome expression (West et al., 1977; Frels et al., 1979; Frels and Chapman, 1980; Papaioannou et al., 1981; Takagi et al., 1982).

Non-random inactivation of the paternal X chromosome has also been observed in the extraembryonic lineages in rats and humans (Wake et al., 1976; Rogers et al., 1978).

The sequences of change in replication pattern, the non-random inactivation in extraembryonic lineages and the behaviour of the HPRT gene from the inactive paternal X chromosome of primitive endoderm in DNA mediated gene transfer experiments (Kratzer et al., 1983; also see section 1.11.1.2), suggested to some investigators (Takagi et al., 1982; Gartler and Riggs, 1983) that random and non-random X inactivation may be two different processes.

However, even the random process of X inactivation in embryonic lineages is under genetic control, since different X-linked mutants are associated with different probabilities of inactivation. The evidence for the existence of a site responsible for controlling the X inactivation process came from classic genetic analysis on the mouse (Cattanach and Isaacson, 1967; Cattanach et al., 1969; 1970; Cattanach, 1970; Cattanach and Williams, 1972). This site, known as the X-chromosome controlling element (Xce), has been mapped to the centre of the linkage group close to Ta (Cattanach et al., 1970). Ta is a coat colour marker situated approximately 3 mu from the mouse pgk-1 gene (Cattanach and Papworth, 1981). There are three known alleles at Xce which act in a cis dominant manner to determine the probability of an X chromosome containing them becoming inactivated.

Cattanach and Isaacson (1967) first identified different Xce alleles in studies involving X-autosome translocations. They showed that these alleles directed the coat colour variegation caused by the translocation. It was later shown that these Xce alleles also

directed the heterozygous phenotypes of two X-linked genes (Cattanach et al., 1969; Cattanach, 1975). Others (Grahn et al., 1970; Krzanowska and Wabik, 1971; Falconer and Isaacson, 1972; Ohno et al., 1973) working with various gene markers have provided further evidence of an X locus controlling the heterozygous phenotypes of X-linked genes. In all these cases, the X controlling locus maps to the same location suggesting that it is a single element (Cattanach, 1975). The Xce alleles do not change dramatically the randomness of X inactivation. The variegation may extend from 50:50 in homozygotes to 70:30 in the Xce^a/Xce^c heterozygotes. This has been nicely demonstrated in mice of pgk-1^a Xce^c/pgk-1^b Xce^a genotype (Cattanach et al., 1983).

1.6 Single X inactivation centre (Xic): the spreading effect:

Genes of the same X chromosome are activated or inactivated together in a cell. As is previously discussed, studies have confirmed that X-linked markers are active on both chromosomes prior to the inactivation. Thus the first event is the initiation of the inactivation of the entire X chromosome. Chromosomal rearrangements between the autosomes and the X chromosome are valuable for studying the mechanisms of this event and to search for controlling centers. When the rearranged X is inactivated during the course of development, all the loci lying close to the heterochromatin are likely to be inactivated. This is called the spreading effect (Lewis, 1950; Baker, 1964; Spofford, 1975).

The evidence from X-autosome translocation studies in the mouse

was considered to support, by some investigators, the concept of one inactivation centre on the X chromosome (Russel and Cacheiro, 1978; Gartler and Riggs, 1983) and by others the concept of multiple X inactivation centers (Eicher, 1970; Disteche et al., 1981; Nakagome, 1982).

Cytogenetic data from several reciprocal X autosome translocations in the mouse indicated that the pieces of the translocated X chromosome were never late labelling in any cell whereas the intact X was inactivated (Russell and Cacheiro, 1978; Rastan, 1983). This pattern of non-random X inactivation could result from initial random X inactivation followed by selection eliminating unbalanced cells. The observation that the two pieces of a reciprocally translocated X chromosome cannot be inactivated was considered to support a model based on the concept of a single inactivation center from which inactivation spreads in cis in both directions.

In Cattanach's translocations (T(X;7) 1 Ct), a piece of the autosomal chromosome 7 carrying several markers was inserted into the X chromosome. The inactivation of both parts of the divided X chromosomes along with the appearance of a non-inactivated region in the autosomal insert was used as evidence to support the model of multiple inactivation centers (Eicher, 1970). This concept assumes that each autosomal insert into the X chromosome should interrupt the spreading of inactivation. However, we know that not every segment on the inactive X chromosome is inactivated, suggesting that inactivation may "skip" regions lacking the proper signals.

To clarify this controversy, Searle's murine X-autosome reciprocal translocation (T (X;16) 16 H) is very valuable since it is a

translocation between the X and chromosome 16 that has the breakpoint in the middle of chromosome X. Female cells carry two normal chromosomes: X and 16, and two rearranged chromosomes, X^{16}, which consists of the proximal part of the X and the distal part of 16; and 16^{X}, which consists of the proximal part of the 16 and the distal part of the X. Searle's translocation is characterized by a non-random inactivation pattern. In heterozygous females the intact X is always inactive (Lyon, 1966; Lyon et al., 1964). Takagi (1980) studied embryos at different stages of gestation (6.5 - 8.5 days gestation) and identified the inactive X by using 5-Bromo-2'-deoxyuridine labelling. He found that most cells at all studied stages had the intact X as late replicating. In 6.5 day embryos, few cells showed late replicating 16^{X}. The X^{16} was never inactivated in any of the almost 2500 embryonic cells examined. Takagi (1980) also studied 322 cells in 7.5 day unbalanced embryos of the karyotype 40 (X/X^{16}, 16/16). Despite the fact that inactivation of the X^{16} chromosome would have led to genetic balance he did not find any asynchronously replicating chromosome. Therefore X^{16} cannot be inactivated due to lack of an inactivation centre in the proximal region of mouse X chromosome, whereas 16^{X} can be inactivated then eliminated by selection for cells with the maximum genetic balance (McMahon and Monk, 1983).

Rastan and Robertson (1985) reported results supporting a single inactivation centre. The prediction of this model is that if Xic has been deleted, neither the X nor the deleted X present in a diploid cell can be inactivated. They have studied six different embryo-derived (EK) cell lines, each with a partial deletion of the distal

piece of one of the X chromosomes. Her cytogenetic results confirmed the above predictions.

Studies on X-autosome translocations in man (Mohandas et al., 1987) and on human X chromosome abnormalities (Therman and Patau, 1975; Summitt et al., 1978; Mattei et al., 1982) also supported the model of a single Xic.

In summary, inactivation spreads in cis in both directions along the X-chromosome. Studies on X-autosome translocations indicated that inactivation may or may not spread into the adjoining euchromatin, or may spread into distal regions while leaving proximal regions non-suppressed (Keitges and Palmer, 1986).

Studies on X-autosome translocations in mice have mapped the Xic to the same region as the Xce (Cattanach and Papworth, 1981; Rastan, 1983). Thus the site of initiation of X inactivation may represent the site of alleles which cause primary non-randomness of X inactivation (see section 1.5).

The evolution of X chromosome inactivation has been conserved in mammals and if these 2 loci are identical in mice, they also should be identical in man. In fact, the Xic in human has been located in the long arm of the X chromosome (at band Xq21) near the pgk-1 locus (Therman et al., 1974; Therman et al., 1979; Francke and Taggart, 1980). Furthermore the location of the Xic appears to be similar in most primates and in some mammals (Flejter et al., 1986).

1.7 The role of autosomal gene(s) in X chromosome inactivation:

Only one X chromosome remains active in a diploid mammalian cell

with extra sex chromosomes generated by meiotic and mitotic errors (Jacobs and Strong, 1959; Ferguson-Smith et al., 1960; Carr and Barr, 1961; Miller, 1961; Ohno, 1969). For example, two X chromosomes are active in a tetraploid cell. The number of active X chromosomes could be one or two in triploids (Weaver et al., 1975; Migeon et al., 1981). These studies suggested that for each diploid set of autosomes there is one active X chromosome. Thus, autosomal factors may be involved in this counting mechanism.

1.8 The role of DNA methylation in X chromosome imprinting:

The non-random inactivation in marsupials and in extraembryonic lineages of eutherian mammals (see section 1.5) indicated that the paternal X is imprinted before or soon after fertilization (Chandra and Brown, 1975). It is useful to look at the germ cell nuclei for DNA modifications that could be responsible for the preferential inactivation of the paternal X chromosome. I have described in section 1.1 the X chromosome inactivation in female germ line and the occurrence of reactivation during oogenesis. During spermatogenesis, the mammalian X chromosome becomes transcriptionally inactive at the onset of the meiotic prophase (Lifschytz and Lindsley, 1974). DNA mediated transfer experiments have suggested that X inactivation in extraembryonic lineages and in spermatocytes appears to be different at the molecular level from inactivation occurring in somatic cells. As described in section 1.15.1.2, the hprt genes in sperm DNA and on the paternal inactive X chromosome of yolk sac endoderm function in gene transfection experiments as well as the active gene. These observations led some to search for a difference in methylation

between the sperm and oocytes gamete genomes and they showed that both genomes are equally undermethylated in minor satellite sequences but that MIF sequences are more methylated in sperm than in oocyte DNA (Sanford et al., 1984; 1988). By day 7 the satellite sequences become methylated in the embryonic lineages (Sanford et al., 1985). In contrast, the opposite is seen in extraembryonic lineages. The repetitive DNA sequences undergo almost total demethylation. Thus, there appears to be a correlation between random inactivation and de novo methylation in the cells of the embryonic ectoderm. However, for methylation to be a plausible model for imprinting, it is necessary to study the differences in DNA methylation of X-linked sequences of oocyte and sperm DNA.

1.9 The expression of genes inserted on the X chromosome:

X-autosome translocation studies showed that the inactivation may or may not spread into the autosome (see section 1.6). Investigators have generated transgenic mice having the transgene inserted on the X chromosome. In the case of transgenic mice harboring the complete chicken transferrin gene, the inserted gene on the inactive X chromosome escaped X chromosome inactivation (1987). Krumlauf et al. (1986) generated transgenic mice carrying the mouse a-fetoprotein (AFP) minigene. Their results indicated that the AFP minigene is expressed on the inactive X chromosome in the visceral endoderm but not in the liver. The inability of these two transgenes to undergo X chromosome inactivation may be due to the lack of regulatory sequences specific to the X chromosome which are necessary for inactivation.

Both genes, the chicken transferrin gene and the AFP minigene, are normally autosomal and one might not expect to find associated with them sequences important for their inactivation. To clarify this issue, it is essential to introduce X-linked genes onto the X chromosome and to study their behaviour on the inactive X chromosome. Will these genes be correctly inactivated?

1.10 The activity of the X chromosomes in embryonal carcinoma (EC) cells:

1.10.1 EC cells: the ideal system to investigate X chromosome inactivation:

X-chromosome inactivation is an intriguing system of control of gene expression. Inactivation starts to occur at the 16-cell morula stage, that is when the embryo consists of a small number of cells and is therefore difficult to analyze experimentally. In addition, X inactivation initiates early but it is not completed until day 6.5 in mouse embryogenesis. Thus the asynchrony of this process makes its study in the embryos a great challenge. These problems could be circumvented by using EC cells as the system to reveal the details of the X inactivation mechanism.

EC cells are the stem cells of malignant tumours known as teratocarcinomas (Stevens, 1967; Pierce, 1967). They resemble the pluripotent cells of inner cell mass since under appropriate conditions in culture they undergo a process of differentiation culminating in the generation of all the cell types normally present

in the embryo (Kahan and Ephrussi, 1970; Martin and Evans, 1975; Nicholas et al., 1975; McBurney, 1976). In addition, EC cells and embryonic cells share several morphological (Pierce and Beals, 1964; Lo and Gilula, 1980;) antigenic (Artzt et al , 1973; Jacob, 1977; Reisner et al., 1977; Solter and Knowles 1978; Kermeler et al., 1977; Boller and Kemler, 1982; Harris et al., 1986) and biochemical (Damjanov et al., 1971; Bernstein et al., 1973) characteristics. That EC cells are indeed similar to the pluripotent embryonic cells have been confirmed by blastocyst injection experiments. EC cells can contribute to the formation of a normal mouse (Brinster, 1974; Papaioannou et al., 1976; Dewey and Martin, 1980; Rossant and McBurney, 1982)

1.10.2 The activity of two X-chromosomes in undifferentiated EC cells:

Several sources of data confirmed the similarity between EC cells and the embryonic inner cell mass. Thus it would be expected that both X chromosomes would be active in undifferentiated female EC cells and that one would be inactivated when cells are differentiated in culture (McBurney and Adamson, 1976; Martin et al., 1978; McBurney and Strutt, 1980).

Both cytogenetic and biochemical data indicated that EC cells may carry two active X chromosomes. McBurney and Strutt (1980) developed a female EC cell line called P10, which is heterozygous for pgk-1 ($pgk\text{-}1^{a}/pgk\text{-}1^{b}$). All clonal populations of these cells contained almost equal levels of both PGK-1 isozymes suggesting that these cells

have 2 active X chromosomes. This suggestion was also strengthened by cytogenetic studies revealing that P10 cells possessed two synchronously replicating X chromosomes. Martin et al. (1978), Takagi and Martin (1984) described similar results for an EC cell line, LT-1, derived from a spontaneous ovarian teratocarcinoma. That LT-1 does indeed carry two active X chromosomes was demonstrated by comparisons of X encoded enzyme activities in LT-1 cells and three undifferentiated XO EC cell lines, which showed that 2:1 dosage relationship expected if LT-1 cells contained two active X chromosomes (Martin et al., 1978).

1.10.2.1 Heterogeneity of X inactivation in undifferentiated EC cells:

Dr. McBurney's laboratory developed and characterized four EC cell lines listed in the table below. Each has unique X chromosome characteristics.

Female EC Cell Line	Number of Active X-Chromosomes	Number of Late Replicating X-Chromosomes	Reactivation with 5-azacytidine
P10	2	0	
C100AG1(XX)	1	0	+
C86S1A1(XX)	1	1	+
C145fAG11(XX)	1	1	-

The three EC cell lines C86S1A1(XX), C100AG1(XX) and C145fAG11(XX) are 8-azaguanine-resistant clones. The frequency of recovery of HPRT-deficient ($HPRT^-$) from the parental cells suggested the presence of an inactive X chromosome in each of these cells (McBurney and Adamson, 1976). Another evidence for X inactivation

comes from cytogenetic analysis demonstrating the presence of an allocyclic X chromosome in C86S1A1(XX) and C145fAG11(XX). However, allocyclic replication of X chromosomes was not present in C100AG1(XX) indicating that this cell line may be arrested at an intermediate stage of X chromosome differentiation. The status of the X chromosomes in these female EC cell lines and the fact that some of these cell lines respond to 5-azacytidine by reactivating their inactive X chromosome support the view that these four lines from P10 to C145 are frozen at different stages of X chromosome differentiation (Paterno et al., 1985; Hockey, Adra and McBurney, manuscript in preparation).

1.10.2.2 Rapid and synchronous X inactivation during differentiation of EC cells:

EC cells can be induced to differentiate in vitro by changes in culture conditions such as the addition of retinoic acid (Martin et al., 1978; McBurney and Strutt, 1980; Jones-Villeneuve et al., 1982; Takagi and Martin, 1984). Martin et al. (1978) reported 50% reduction in the dosage of X chromosome gene products in differentiated XXLT-1 cells. McBurney and Strutt (1980) as well as Takagi and Martin (1984) demonstrated the appearance of allocyclic X chromosome in the majority of differentiated cells. Both EC cell lines, P10 and C86 (see section 1.10.2.1) can be induced to differentiate following exposure to retinoic acid and the differentiation of P10 is accompanied by the inactivation of 1 X-chromosome in each cell (McBurney and Strutt, 1980; Paterno et al., 1985; Paterno, 1985, Ph.D. Thesis).

Thus, the major properties of dosage compensation that occur in embryological development occur also in EC cells. This as well as the

fact that these cells are more amenable to experimentation make the EC cells an experimentally valuable system for studying the regulation of X chromosome differentiation during development.

1.11 Molecular structure of genes:

In 1977, it was discovered that eukaryotic genes such as immunoglobulin (Brack and Tonegawa, 1977), globin (Jeffreys and Flavell, 1977) and tRNA (Goodman et al., 1977) genes are interrupted by insertions of non-coding sequences. These sequences which are called introns are spliced out of the precursor RNA to generate a functional mRNA. The DNA sequences which are represented in the mature mRNA are called exons. Now that a large number of genes have been sequenced it appears that the majority contain insertions of non-coding DNA. Some genes such as the histone genes have no introns.

Introns are present in mammalian genes, are rare in those of lower eukaryotes and are absent from those of eubacteria (Cornish-Bowden, 1985). Several archaebacteria and bacteriophage T4 genes contain introns. Unlike exons, which are usually short (40-50 amino acids), introns can be very large (Gilbert, 1985; Koenig et al., 1987) thus accounting for the large size of some vertebrate genes.

The positions of the introns relative to the mRNA sequence of many genes are conserved in different mammalian species (Blake, 1983; Flemington et al., 1987; Traut, 1988).

1.12 Promoters of genes that encode proteins:

The first level of control of gene expression in eukaryotic cells

is in the initiation of transcription. RNA polymerase II initiates transcription by recognizing cis-acting sequences called promoters. Promoters are located upstream from the start site of various genes. A comparison of several gene promoter regions and a detailed functional analysis revealed the presence of conserved short stretches in a typical promoter: the start site, a region centered at about -25 and one or more regions of 8 to 12 base pairs (bp) named upstream promoter elements (UPEs) (for review see Dynan and Tjian, 1985; McKnight and Tjian, 1986; Maniatis et al., 1987).

The start site is the position on DNA that corresponds to the first base of the transcript. A consensus sequence has emerged by aligning the sequences of a large number of promoters (Corden et al., 1980):

5'-PyCAPyPyPyPyPy-3'

The region which is typically located at position -25 is designated as the TATA or Goldberg-Hogness box (Goldberg, 1979). It has the consensus sequence: $TATA\frac{A}{T}A\frac{A}{T}$ (Breathnach and Chambon, 1981). This sequence determines the site of transcription initiation. In the absence of the TATA consensus sequence, transcription initiates at multiple start sites (Corden et al., 1980; Reynolds et al., 1984; Melton et al., 1986; Ingolia et al., 1986; Flemington et al., 1987).

The so-called UPEs such as the CAAT box (Efstratiadis et al., 1980) and the GC box (Benoit and Chambon, 1981; Myers et al., 1981; Everett et al., 1983) are located farther upstream at about 40-100 nucleotides upstream of the start site (reviewed by Dynan and Tjian, 1985; McKnight and Tjian, 1986). Promoter mapping studies have shown that UPEs often function to stimulate the level of transcription

regardless of their orientation with respect to the TATA box (Maniatis et al., 1987).

DNA footprinting (Galas and Schmitz, 1978) and gel retardation (Fried and Crothers, 1981) studies have revealed the presence of UPE-binding factors (reviewed by McKnight and Tjian, 1986). Two of these factors have been extensively purified and studies: Sp1 which interacts with the GC box (reviewed by Kadonaga et al., 1986) and CTF which recognizes the CAAT box (Jones et al., 1985). Both factors are proteins and activate RNA polymerase II transcription.

1.13 Promotors of X-chromosomal genes

Regulatory sequences which might be unique to the X-chromosomal genes could be revealed by sequence analysis. Now, several X-linked genes have been sequenced. Examples include: the human factor VIII (Gitschier et al., 1984) and factor IX (Yoshitake et al., 1985), the human pgk-1 (Singer-Sam, 1984), the human g6pd (Martini et al., 1986) and the human and mouse hprt (Melton et al., 1986; Patel et al., 1986). A comparison of their promoter regions (done by Dr. M.W. McBurney) failed to detect sequences which are unique features of X-chromosomal genes. However we cannot exclude the possibility that regulatory sequences may be found in regions which have not yet been examined.

The major features of the promoters of the constitutively expressed X-linked genes are similar to those of a variety of autosomal housekeeping genes such as hmg CoA reductase (Reynolds et al., 1984), ada (Ingolia et al., 1986), aprt (Broderick et al., 1987) and tk (Flemington et al., 1987). All of these genes have multiple GC

boxes within their promoters. They lack TATA boxes, and may or may not have a CAAT box. Unlike some of the promoters of tissue-specific genes (Cullen et al., 1986), all of these promoters have clusters of CpG dinucleotides which are potential sites for DNA methylation (see section 1.5).

1.14 Retroposons or processed pseudogenes:

The mammalian genome contains large numbers of processed pseudogenes (Vanin, 1985; Rogers, 1985). Each processed pseudogene is thought to have arisen by RNA-mediated duplication from a functional gene because pseudogenes generally lack introns, contain poly A tracts at their 3' ends, and are flanked by short direct repeats of 7-21 nt. In general, the processed pseudogenes or retroposons are not genetically linked to the parental gene and are not expressed because they have no promoter and because their sequences contain termination codons in all reading frames. Processed pseudogenes are abundant in mammals, but not other organisms, and they are thought to have evolved relatively recently, within the past 30 million years (Weiner, 1985).

1.15 DNA methylation:

It was discovered more than 40 years ago that 5-methylcytosine occurs as a minor base in the DNA of many organisms (Hotchkiss, 1948; Wyatt, 1951; Dunn and Smith, 1958). The addition of a methyl group to the number 5 carbon of some cytosine residues is a post-replicative process (Gold et al., 1963; Kalousek and Morris, 1968). The

methylated cytosine residue is usually adjacent to a guanine residue on the 3' side (Doskocil and Sorm, 1962; Salomon and Kay, 1970; Sneider, 1972; Gerlach and Bedbrook, 1979) and the cytosine residue base paired with that guanine is also methylated (Bird, 1978). Several models have been proposed associating DNA methylation with differential activity of genes in different tissues during development (Scarano, 1971; Holliday and Pugh, 1975; Sager and Kitchin, 1975) and with X inactivation (Riggs, 1975; Sager and Kitchin, 1975). The essential postulates of Holliday and Pugh (1975) and of Riggs (1975) models were:

a) interaction between proteins and DNA sequences is responsible for gene regulation, X inactivation and differential activity of genes during development;

b) methylation should affect regulatory protein binding; and

c) the most important postulate was that there should be a maintenance DNA methylase (in addition to a <u>de novo</u> methylase) which will convert a half methylated site:

$$5'-{}^{m}CG-3'$$

$$3'-GC-5'$$

to a symmetrically methylated site:

$$5'-{}^{m}CG-3'$$

$$3'-GC_{m}-5'$$

thus, once methylation events occur after fertilization, the methylated sites are inherited through DNA replication but not transmitted through the germ line.

It is of interest to note that these models were proposed without any direct evidence for any eukaryotic specific modification enzymes

and for any function of the methylation of eukaryotic DNA.

In recent years, investigators have accumulated evidence in favor of these models.

By inserting Moloney leukemia virus genomes into mouse strains Jähner et al. (1982) showed that de novo methylation occurred when the virus was introduced in preimplantation but not into postimplantation embryos. De novo methylation of retroviruses was also observed in EC cells (Stewart et al., 1982). This activity observed with foreign DNA reflects the naturally occurring processes since cellular sequences also undergo methylation during embryonic development (Sturm and Taylor, 1981; Pages and Roizes, 1982; Ehrlich et al., 1982; Adams et al., 1983; Chapman et al., 1984).

Once the methylation pattern is formed, it is inherited through DNA replication by a maintenance methylase which is involved in methylating hemimethylated DNA (Gruenbaum et al., 1982). However, it is likely that de novo activity and maintenance activity are carried out by the same protein (Bestor and Ingram, 1983; Sano et al., 1983; Pfeifer et al., 1983; Razin and Szyf, 1984).

One line of evidence is consistent with the view that methylation may be involved in X inactivation. In vivo studies relied on the two restriction endonucleases HpaII and MspI to correlate gene activity and methylation. For a large number of genes, such as the globin genes (Perry, 1983), the chicken ovalbumin (Mandel and Chambon, 1979) and the rat insulin II gene (Cate et al., 1983; Kuo et al., 1979), some CpG dinucleotides became completely unmethylated in the expressing tissue (for review see Razin and Szyf, 1984; Jaenisch and Jähner, 1984). There are several exceptions to this rule. For

example, the S-crystalline gene (Grainger et al., 1983) and the chicken and Xenopus vitellogenin genes (Perry, 1983; Burch and Weintraub, 1983; Gerber-Huber et al., 1983) are highly methylated yet active. However, the value of the latter results is questionable since the investigators have relied on two enzymes: MspI and HpaII which detect only 10-15% of the total CpGs.

These experiments do not allow the establishment of a cause and effect relationship between gene inactivity and methylation since methylation could be a secondary effect of transcriptional repression. To test for causative relationship, the approach was to methylate in vitro some genes, transfer them into cells and analyze their expression in a functional assay. Experiments using this strategy demonstrated that in vitro methylation of HpaII sites abolished the activity of certain cloned cellular (Wigler et al., 1981; Stein et al., 1982) and viral genes (Simon et al., 1983; Kruczek and Doerfler, 1983; Fradin et al., 1982; Jaenisch and Jähner, 1984). In addition, treatment with demethylating agents such as 5-azacytidine resulted in their activation (Jones, 1984). The critical sites relevant for gene activity are usually localized at the 5' end in the regulatory region of a gene (Kruczek and Doerfler, 1983; Busslinger et al., 1983). These experiments suggested a causative relationship between gene expression and methylation.

1.15.1 DNA methylation and X chromosome inactivation:

1.15.1.1 Reactivation by treatment with demethylating agents:

Mohandas et al. (1981) demonstrated that 5-azacytidine treatment

of a hypoxanthine phosphoribosyl transferase (HPRT) deficient mouse-human cell hybrid carrying a wild type hprt allele on the inactive human X chromosome elevated by 1,000 fold the frequency of expression of this allele thus enabling the cell to form HAT resistant colonies. In addition, two unselected human X-linked markers, glucose 6 phosphate dehydrogenase (G6PD) and phosphoglycerate kinase (pgk-1), were expressed in a small proportion of the $HPRT^+$ cells. Although the reactivation was not a whole chromosome event, this was the first report of reactivation of an X-linked gene by treatment with a hypomethylating agent suggesting that DNA methylation may play a role in X inactivation. Similar studies using HPRT-deficient human-mouse hybrid (Lester et al., 1982; Hors-Cayla et al., 1983) and HPRT-deficient M. musculus X M. caroli lines (Graves, 1982) confirmed that 5-azacytidine treatment increased the appearance of $HPRT^+$ colonies and that occasionally reactivants expressing all four human markers, g6pd, pgk-1, alpha-gal and hprt could be recovered (Hors-Cayla et al., 1983; Graves, 1983). It is important to note that in all these cases the reactivation was not a whole chromosome event and that the reactivants continued to carry heterochromatic late replicating human X chromosomes.

Recently we (Paterno et al., 1985) reported our experiments with a HPRT-deficient EC cell line, named C86S1A1(XX), which contains an inactive X chromosome. The treatment of these cells with 5-azacytidine caused the stable reactivation of the $HPRT^+$ allele in 10% of the cells. Furthermore, the HAT resistant clones replicated both X chromosomes synchronously with the autosomes (Paterno et al., 1985; Hockey, A.G., Adra, C.N. and McBurney, M.W., manuscript in

preparation). This total reversal of the heterochromatic pattern of the inactive X was not observed in the above described cases. Following X chromosome reactivation by treatment with 5-azacytidine, reactivation is extremely stable even without selection. Loss of $HPRT^+$ activity was always associated with the loss of the X chromosome carrying the reactivated HPRT allele (Grave, 1982; 1983; Lester et al., 1982; Mohandas et al., 1981; Paterno et al., 1985).

The most important conclusions from the above reactivation studies are:

a) reactivation involves a stable and heritable change in gene activity suggesting that the spreading of inactivation does not re-occur more than once in the life of a cell;

b) the reversibility of the inactivation, as in germ cells, implies that major DNA rearrangement is not involved in X inactivation; and

c) 5-azacytidine reactivation of X-linked genes supports the hypothesis of Riggs (1975) and of Holliday and Pugh (1975) that DNA methylation plays a role in X inactivation.

1.15.1.2 DNA mediated gene transfer experiments of X chromosome activity:

Further evidence for the role of DNA methylation in the maintenance of X inactivation came from DNA transfection experiments. Liskay and Evans (1980) were the first to show that the hprt gene on the inactive X chromosome does not function in DNA-mediated gene transfer into a recipient cell. Chapman et al. (1982) used DNA from

mouse cells heterozygous for electrophoretic variants of HPRT and Venolia and Gartler (1983) used DNA from human cells heterozygous for a Lesch-Nyhan $HPRT^-$ mutation in similar experiments. Both laboratories confirmed that the hprt gene on the inactive X chromosome does not function in DNA transfection experiments. DeJonge et al. (1982) reported surprising results claiming the contrary. However, the significance of their observation is questionable since it could not be repeated by others (Venolia and Gartler, 1983). Thus all these experiments suggest that genes on the inactive X chromosome are not able to be expressed following transfection. Since in vitro methylation of some genes abolishes their expression after transfection (see section 1.15), the lack of HPRT activity implies that the hprt gene may be methylated.

This hypothesis was tested by using DNA from a mouse-human cell hybrid containing the human inactive X chromosome and from $HPRT^+$ reactivants (Mohandas et al., 1982) in DNA transfection experiments (Venolia et al., 1982; Lester et al., 1982). Only the DNA from the 5-azacytidine induced reactivants did function. Analysis of the 5' region of the hprt gene revealed the presence of a cluster of CCGG sites that is hypomethylated on the active X chromosome (Xa) (Yen et al., 1984; Wolf et al., 1984).

It is of interest to note that the hprt gene in sperm DNA (Venolia et al., 1984) does function in DNA-mediated gene transfer in spite of the fact that the X chromosome is inactive during spermatogenesis. In addition, the inactive hprt gene carried on the paternal inactive X chromosome of yolk sac endoderm functions in gene transfection as well as the active gene (Kratzer et al., 1983). These

results suggest that the mechanism of X inactivation in extraembryonic and spermatogonial cells appears to be different from the one involved in X inactivation in somatic lineages.

In summary, the evidence is strong to support the claim that DNA methylation is involved in controlling the activity of X-linked genes at least in female somatic cells. But what sort of control could that be? In answering that question, we should look into experiments which showed that there is no correlation between methylation and gene activity. Treatment with 5-azacytidine did not induce reactiva- tion of the inactive X in normal human diploid cells (Wolf and Migeon, 1982) and in some EC cell lines (Paterno et al., 1985). This supports the notion of methylation being a secondary controller of gene expression, a controller that acts to stabilize the inactive state.

1.15.1.3 Methylation pattern of X-linked genes and DNA sequences:

To investigate the molecular mechanisms responsible for X inactivation, it is necessary to use cloned X-specific sequences and methylation sensitive restriction endonucleases to construct restriction maps of the cloned region on the active and inactive X chromosome (Methylation studies of non X-linked sequences were discussed earlier in section 1.15.1). Using this approach researchers evaluated the degree of involvement of DNA methylation in X chromosome inactivation.

Wolf and Migeon (1982) using two single-copy X chromosome-specific DNA sequences which map to the long arm of the human X, found no differences in DNA methylation patterns between the active and

inactive X chromosomes. However, their study has some weakness. The authors relied on the MspI/HpaII enzymes which measures only 15% of the total CpGs. These sites may not be relevant for gene activity. Moreover, since not all loci on the inactive X chromosome are subject to inactivation, the probes used by Wolf and Migeon may not be part of genes that undergo X inactivation.

Other studies of the human (Yen et al., 1984; Wolf et al., 1984) and mouse (Lock et al., 1986) hprt genes, the human g6pd gene (Toniolo et al., 1984; Wolf et al., 1984) and the human pgk-1 gene (Keith et al., 1986) showed that certain regions of these genes are hypermethylated on the inactive X chromosome (Xi).

To establish whether the methylation occurs before or after X inactivation, Lock and her co-workers (1987) examined the methylation pattern of the mouse hprt gene in embryonal carcinoma cells and in female embryos at various stages of embryogenesis. They observed that X inactivation preceded the methylation of the hprt gene. These data suggested that DNA methylation is a secondary mechanism by which the cell maintain the X chromosome in its inactive state. However, Lock and her co-workers studied only the upstream region of the hprt gene and it is likely that this region may not be relevant to the mechanism of X inactivation (Cullen et al., 1986). In addition, only a subset of the CpG sequences present in the 5' end of the hprt gene can be recognized by the available enzymes.

1.16 Aims of the thesis:

In this thesis some aspects of the molecular basis of X

chromosome inactivation have been investigated. Also, the cloning and the characterization of the mouse pgk-1 gene family has been described.

Because of strong evidence for some role of DNA methylation in X inactivation and because of the fact that X inactivation can be analyzed in culture using female EC cell lines, we decided to examine the molecular mechanisms involved in X inactivation by using DNA demethylating agents to reactivate the inactive X chromosomes in our EC cell lines (Paterno et al., 1985; also see Appendix 1) and by cloning and characterizing the mouse pgk-1 gene (Chapter III). Our earlier attempts to clone the pgk-1 cDNA using synthetic oligonucleotides led to the isolation of a single clone. This clone was sequenced and shown not to encode PGK-1 (Appendix 3).

To test for a causative relationship between gene inactivity and methylation, I investigated the methylation pattern at HpaII and HhaI sites present in the 5' control region of the pgk-1 gene of kidney, spleen and EC cell lines before and after differentiation (Chapter IV).

The mammalian genome contains two genes encoding phosphoglycerate kinase. The pgk-1 is X-linked and is expressed in all somatic cells and in premeiotic germ cells, while the pgk-2 gene is expressed in postmeiotic germ cells, where the pgk-1 gene is turned off. To understand the evolutionary relationship between these two genes and to set the stage for the investigation of the mechanism responsible for their regulation in spermatogenic cells, I decided to clone the pgk-2 gene (Chapter V). My studies also have revealed the presence of other pgk-1-related sequences (Chapter V).

CHAPTER TWO

MATERIALS AND METHODS

2.1 Cell lines and culture conditions

The cell lines used in the experiments are listed in Table 2.1 with some of their properties. They are all derived from teratocarcinomas generated in C3H/He mice by transplantation of embryos. All cell lines were cultured in alpha minimal essential medium (Stanners et al., 1971) (Gibco Laboratories, Grand Island, N.Y.), containing 7.5% calf serum (Animal Health Laboratories Inc., Toronto, Canada), 2.5% fetal calf serum (Gibco Laboratories, Chagrin Falls, Ohio) or Clex serum (Dextran Products Limited, Scarborough, Ontario), penicillin (50 mg/ml) and 100 mg/ml streptomycin. They were maintained in Falcon plastic tissue culture dishes (Becton, Dickenson and Co., Oxnard, CA, U.S.A.), placed at 37°C in a 5% CO_2 incubator. The cells were routinely subcultured every 2 days by washing them first with Ca^{++}- and Mg^{++}-free phosphate buffered saline (PBS) then incubating them with 0.025% trypsin and 1 mM ethylenediamine tetraacetic acid in PBS for 5 min in order to detach them from the surface of the dish. The cells were resuspended by pipetting them up and down several times, then counted and replated at a concentration of 10^5 cells per ml into fresh medium.

TABLE 2.1

Properties of the Embryonal Carcinoma Cells

Cell Line	Sex Chromosomes	Late-Replicating X-Chromosomes	Genotype	5-Azacytine Reactivatable X-Chromosomes
P10	XX	0	$pgk\text{-}1^{a},Xce^{c}/pgk\text{-}1^{b},Xce^{a}$	NA
P10(XO)	XO	0	$pgk\text{-}1^{a},Xce^{c}$ or $pgk\text{-}1^{b},Xce^{a}$	NA
C86S1	XX	1	$Xa,hprt^{+}/Xi,hprt^{+}$	YES
AGM2	XX	1	$Xa,hprt^{s}/Xi,hprt^{+}$ derived from C8S1	YES
C86S1A1	XX	1	$Xa,hprt^{-}/Xi,hprt^{+}$ derived from C86S1	YES
C86S1A1AZA(n)	XX	0	derived from C86S1A1 by treatment with 5-AC	NA
$C86S1A1_{n}$(XO)	XO	0	$Xa,hprt^{-}$ derived from C86S1A1	NA
C100AG1(XX)	XX	0	$Xa,hprt^{-}/Xi,hprt^{+}$	YES
C145fAG11(XX)	XX	1	$Xa,hprt^{-}/Xi,hprt^{+}$	NO

$hprt^{s}$, a mutant hprt allele which encodes a thermolabile HPRT

n, indicates a number of clones derived from the parental cell line

NA, not applicable

2.2 RNA preparations

2.2.1 Extraction of total RNA from cell lines

RNA was purified by modifying the protocol of Auffray and Rougeon (1980). Cells were cultured until they reached 5 x 10^7 cells/plate. The medium was aspirated and the cells were washed three times with 10 ml of ice-cold PBS, then 2 ml of 8 M urea/3 M LiCl (made up in H_2O treated with 0.1% diethyl pyrocarbonate (DEPC) overnight and autoclaved to destroy the DEPC) was added to each dish and the cells were scraped and pooled into a 50 ml propylene tube. The DNA of the lysed cells was sheared by a Polytron homogenizer adjusted to speed 7 for three minutes on ice. The RNA was allowed to precipitate by storing the lysate on ice in the cold room for 8 hours. The lysate was transferred to a ribonuclease (RNAse) free tube (tube was pre-rinsed with 10% SDS, washed with DEPC H_2O then autoclaved) and the RNA collected by centrifugation at 4°C, in a JA-20 rotor at 6000 rpm for 10 min. The pellet was washed twice with 3 M LiCl (made up in DEPC H_2O then filtered to remove impurities) with centrifugation after each rinse. Finally the RNA was dissolved in DEPC H_2O and stored at -70°C. Approximately 1 mg of RNA was recovered from each plate.

2.2.2 Extraction of total RNA from mouse organs

Mice were dissected and various organs were collected in dishes placed on ice. One g of tissue was transferred to a 50 ml propylene tube containing 10 ml of 3 M LiCl/8 M urea and the preparation of RNA

was carried out as described in section 2.2.1.

2.3 DNA preparation

2.3.1 Genomic DNA preparation

Genomic DNA was prepared using the modified procedure of Blin and Stafford (1976). Five confluent plates of cells ($5x10^6$ cells per plate) were rinsed three times with ice-cold PBS and 2 ml of solution A (10 mM Tris-HCl pH 8.0, 10 mM EDTA, 10 mM NaCl, 0.5% SDS) + 100 mg of proteinase K were added to each plate. Immediately the cells were scraped using a sterile rubber policeman, pooled and poured into a 50 ml polypropylene tube. The tube was incubated at 37°C for 2 hr with gentle shaking. The cell suspension was extracted once with phenol then with chloroform/isoamyl alcohol (24:1 V/V), followed by a treatment with 50 mg/ml pancreatic RNAse and 2 mg/ml RNAse T_1 for 20 min in a shaking 37°C incubator. The sample was then extracted once with phenol, and once with chloroform. The organic phase and aqueous phase were mixed by gentle shaking and separated by centrifugation for 2 min at 3,000 g. The aqueous phase was transferred to a fresh tube between each extraction. To the final aqueous phase 0.1 volume of 3 M sodium acetate and 2 volumes of ice-cold 95% ethanol was added. The solutions were mixed until the precipitate of DNA formed. The DNA, which is visible, was transferred to a microfuge tube, dried and resuspended in the desired volume of TE (0.5 - 2 mg/ml) (TE is 10 mM Tris-HCl and 1 mM EDTA, pH 8.0). 600 mg of DNA were usually recovered from five plates.

2.3.2 Plasmid DNA preparations

Plasmid DNA was extracted from various bacterial clones using the modified protocol of Birnboim and Doly (1979) for rapid small-scale preparations of plasmid DNA and the protocol of M.A. Marko et al. (1982) for large-scale isolation of highly purified plasmid DNA. The method for the rapid, small-scale isolations of plasmid DNA will be described here. Five ml of LB broth (LB is per liter 10 g bacto-tryptone, 5 g bacto-yeast extract, 10 g NaCl, adjusted to pH 7.5 with sodium hydroxide) supplemented with 100 mg/ml ampicillin were inoculated with a single bacterial colony and incubated at 37°C, 300 rpm. for 9 hrs. 1.5 ml of the culture was transferred into an Eppendorf tube and centrifuged for 30 seconds at 12,000 g (the remainder of the culture was stored in the cold room). The medium was removed by aspiration, the pellet was resuspended by vortexing in 100 ml of an ice-cold solution I (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA, 4 mg/ml lysozyme prepared fresh) and then incubated for 3 minutes at 22°C. 200 ml of a freshly prepared solution II (0.2 N NaOH, 1% SDS) was added, the tube was inverted three times for mixing then incubated for 3 min on ice. 150 ml of an ice-cold solution III (5 M potassium acetate, pH 4.8) was added to the cells and mixed by gentle vortexing for 15 seconds. Following an incubation of 3 min on ice, the lysed cells were centrifuged in an Eppendorf centrifuge at 4°C for 5 min. The supernatant was transferred to a fresh Eppendorf tube, extracted once with phenol and once with chloroform/isoamyl alcohol (24:1 V/V). The organic and aqueous phases were mixed by vortexing and separated by centrifugation for 1 min. The aqueous

phase was transferred to a fresh tube between each extraction. To the final aqueous phase two volumes of 95% ethanol were added. Following vortexing, the tube was centrifuged for 5 min at room temperature, the supernatant removed and the DNA pellet dried in a vacuum desiccator. The pellet was resuspended in 50 ml of TE, pH 8.0, containing 20 mg/ml of DNase-free pancreatic RNase. The protocol of M.A. Marko et al. (1982) for large-scale isolation of highly purified plasmid DNA was used without any modifications.

2.4 Phage DNA preparation

The phage DNA preparation was carried out as described by Maniatis et al. (1982).

2.5 Transformation protocols of bacteria

2.5.1 Transformation protocols of E. coli DH1 or DH5

The modified high efficiency technique of Hanahan (1985) was used. A 1 liter Erlenmeyer flask containing 100 ml of SOB medium (2% bacto-tryptone, 0.5% bacto-yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgSO_4$ and 10 mM $MgCl_2$) was inoculated with several E coli DH5 colonies (Hanahan, 1985) off a freshly streaked SOB agar plate. The flask was incubated at 37°C, 200 rpm until the cell density reached 6 x 10^7 cells/ml. The cells were centrifuged in 50 ml sterile polypropylene tubes at 4°C for 15 min at 1000 g. The pellet was resuspended by gentle vortexing in 30 ml of filter sterilized RF1 (RF1 is per

liter 12 g RbCl, 9.9 g Mn Cl_2. $4H_2O$, 30 ml of a 1 M stock (pH 7.5) of potassium acetate, 1.5 g $CaCl_2.2H_2O$, 150 g glycerol, adjusted to pH 5.8 with 0.2 M acetic acid) incubated on ice for 15 min, then recentrifuged again at 1000 g. The bacteria was resuspended in 8 ml of filter sterilized RF2 (RF2 is per liter 10 mM MOPS, 10 mM RbCl, 75 mM, $CaCl_2.2H_2O$, 15% (W/V) glycerol, adjusted to pH 6.8 with NaOH), then incubated on ice for 10 min before freezing 200 ml aliquots in 1.5 ml Eppendorf tubes in liquid nitrogen. The tubes were stored in a freezer at -70°C. Tubes were removed from -70°C and placed at 37°C until just thawed then transferred on ice. The DNA was added in a volume of 50 ml or less. The DNA was mixed with the bacteria by stirring, and the tube was incubated for 25 minutes on ice. The cells were heat shocked in a 42°C water bath for 50 seconds and then returned to 0°C for 5 min. 800 ml of SOC medium (SOC medium is SOB medium + 20 mM glucose) were added to the cells then incubated at 37°C with gentle shaking for 40 min.

Using the above procedure the transformation efficiency was typically 1.5×10^9 transformants/mg DNA.

2.5.2 Preparation of Q359 and P2392 plating bacteria

The preparation of plating bacteria was made by modifying the procedure described by Vector Cloning Systems (San Diego, CA, U.S.A.). A 1 liter Erlenmyer flask containing 100 ml of TB medium (TB medium is per liter 10 g bacto tryptone, 5 g NaCl pH 7.4) supplemented with 0.2% maltose (from a 20% filtered maltose stock) and 10 mM $MgSO_4$ (from a 1 M filtered $MgSO_4$ stock) was inoculated with a colony of a freshly streaked NZY agar plate. The flask was incubated in a 30°C incubator

at 200 rpm for 10 hours. Then the cells were pelleted at 4000 rpm for 5 minutes and resuspended in 40 mls of 10 mM $MgSO_4$.

2.5.3 Plating bacteriophage g

When plating phage onto 100 mm NZY agar plates (NZY medium + 1.5% bacto-agar, NZY medium is per liter 2 g $MgSO_4$-$7H_2O$, 5 g yeast extract, 5 g NaCl, 10 g N-Z-amine (Humko Sheffield Chemical, Norwich, N.Y. 13815), adjust to pH 7.5 with pH 7.5) 0.2 ml of cells adjusted to an optical density at 600 mm of 0.5 with phage dilution buffer (per liter 5.8 g NaCl, 2 g $MgSO_4.7H_2O$, 5 ml 1 M tris-HCl pH 7.5, 5 ml 2% gelatin) were mixed with phage for 20 minutes at 37°C then 3 ml of top NZY agarose were added (NZY medium + 0.7% agarose). The plate was incubated at 37°C for 9 hours.

2.6 Construction of genomic libraries

Lambda libraries were constructed in the vector EMBL3 (Karn et al., 1980; Frischauf et al., 1983) using the Stratagene (San Diego, California 92121) EMBL3 arms and packaging extracts (Kaiser and Murray, 1985). Genomic DNA was prepared from mouse spleen and digested to completion with BamHl. A mixture containing equal moles of vector arms and target DNA was ligated in a final volume of 5 ml using ligase buffer (1 x ligase buffer: 50 mM tris-HCl pH 8.0, 7 mM $MgCl_2$, 1 mM dithiothreitol, and 1 mM ATP) and 200 units of T4 DNA ligase. The reaction was placed at 22°C for 2 hours, then incubated at 4°C for 10 hours. To test the efficiency of the ligation, 0.2 ml

was removed from the ligation reaction, mixed with 10 ml of TE pH 8.0 and heated at 68°C for 10 min. The sample was analyzed by electrophoresis through a 0.6% agarose gel, using as markers g Hind III DNA and an aliquot which was removed from the reaction mixture immediately before the start of the ligation reaction then denatured by heat. 1 mg of the concatamerised DNA was packaged as described by Kaiser and Murray (1985) then 0.5 ml of phage dilution buffer was added (per liter 5.8 g NaCl, 2 g $MgSO_4.7H_2O$, 5 ml 1 M tris-HCl pH 7.5, 5 ml 2% gelatin). To measure any background that might be due to the extracts a mock packaging reaction, containing no DNA, was done. Also to test the efficiency of the extracts, a packaging reaction containing wild-type lambda DNA cI857, concatamerized at a concentration of 210 mg/ml, was also included. The titer of bacteriophages in each of the packaging reactions was measured by preparing two consecutive 10^{-2} dilutions in phage dilution buffer and dispensing 10 ml of each of the 10^{-4} dilution into each of two sterile 10 mm x 50 mm tubes. 0.2 ml of plating bacteria was added to each tube and the tubes were incubated at 37°C with gentle shaking for 20 minutes. 3.0 ml of NZY medium containing melted 0.7% agarose (50°C) were added to each tube, vortexed, and immediately poured onto a labelled warm plate containing 35 ml of bottom NZY agar (NZY medium + 1.5% bacto-agar, NZY medium is per liter 2 g $MgSO_4-7H_2O$, 5 g yeast extract, 5 g NaCl, 10 g N-Z-amine (Humko Sheffield Chemical, Norwich, N.Y. 13815), adjust to pH 7.5 with NaOH). After the hardening of the top agarose the plates were inverted and incubated at 37°C for 9 hours before counting plaques. Uncut wild-type lambda DNA was packaged with efficiencies exceeding 4×10^8 per mg. With no insert, no plaques

were detected on Q359 or P2392. However, the EMBL3 arms when ligated to the target genomic DNA and plated on P2392 or Q359 gave usually 1.5 x 10^6 recombinant plaques per mg.

2.7 Screening of genomic libraries

To screen the genomic library a total of 1 x 10^6 recombinant plaques were examined. Aliquots of the packaging mixture containing 2 x 10^5 infectious bacteriophages were mixed with 2 ml of plating bacteria and incubated at 37°C for 20 min with gentle shaking in a 50 ml sterile tube. 32 ml of NZY medium containing melted 0.7% agarose (50°C) were added to each tube, vortexed and immediately poured onto a 243 x 243 x 18 mm NZY agar dry bio-assay dish (Nunc, Kamstrup, Denmark). The plates were incubated at 37°C until the plaques reached a diameter of 1.0 mm.

Plaques were lifted onto Biodyne transfer membranes (PALL, Glen Cove, New York) and prepared for screening by the technique described in Benton and Davis (1977). Colonies were lifted onto Millipore HATF filters (Millipore Corporation, Bedford, Massachusetts) and prepared for screening by the technique described in Hanahan and Meselson (1980). The libraries were hybridized to the full-length human pgk-1 cDNA (Michelson et al., 1983) in 50 mM Tris-HCl, pH 7.5, 0.3 M NaCl, 50% formamide, 2.5 x Denhardt's solution, 0.5% sodium dodecyl sulfate (SDS), and 250 mg/ml denatured salmon testis DNA (Maniatis et al., 1982). Positive colonies or plaques were rescreened and clones isolated. DNA isolated from each clone was digested with restriction enzymes (usually BamHI) to identify clones carrying fragments

corresponding to those found on genomic blots.

2.8 Subcloning of recombinants in plasmids

1 mg of recombinant DNA or an appropriate quantity allowing me to visualize the fragment of interest was digested by the restriction enzymes under conditions suggested by the manufacturer in a total volume of 20 ml. Similarly, 200 ng of vector DNA (pSP64, pSP65, Gemini 3 or 4) (Melton et al., 1984a) were also digested. After 1 hour of digestion the vector DNA was treated with calf intestinal phosphatase (20 units/1.2 mg DNA, Boehringer, Mannheim, Penzberg) for 20 min at 37°C to prevent self-annealing during ligation. Both samples, along with markers, were electrophoresed in 0.6% low melting agarose gel (Bethesda Research Laboratories, Gaithersburg, MD) in 1 x TBE (0.089 Tris-borate, 0.089 boric acid, 0.002 M EDTA, pH 8.0) at 60 mA at room temperature. After sufficient time had elapsed, the gel was transferred very carefully to a sheet of saran wrap and laid on a black background. Using a long wave U.V. light the bands of interest were visualized and cut with a sharp blade. The agarose pieces were trimmed in order to remove the agarose which does not contain DNA. The fragments were transferred to prelabelled sterile Eppendorf tubes and an equal volume of autoclaved deionized H_2O was added. The agarose was melted at 70°C for 7 minutes, then incubated at 37°C for 5 minutes before initiating the ligation reaction. Only 1/2 of the total volume of each sample was used for the ligation reaction which was accomplished at room temperature for 3 hours in a total incubation volume of 100 ml using ligase buffer (see section 2.6) and 10 units of

T4 DNA ligase. The reaction was terminated by adding 300 ml of sterile H_2O and heating the sample at 65° for 8 min. 50 ml of the ligated product were used to transform competent DH5 cells as described in section 2.5.1.

2.9 DNA analysis

DNA was digested by restriction enzymes under conditions suggested by the manufacturer and electrophoresed in agarose gels in 1 x TPE (0.08 M Tris base, 0.002 M EDTA, 0.08 M phosphoric acid; pH=8.0) containing 0.20 mg/ml ethidium bromide. Genomic and recombinant DNAs were blotted onto Biodyne transfer membranes (PALL). Prehybridization was in 5 x SSPE, 50% formamide, 2.5 x Denhardt's solution, 0.2% SDS, and 0.05 mg/ml denatured salmon testis DNA (1 x SSPE is 0.15 M NaCl, 0.01 M $Na_2H_2PO_4$, 0.001 M EDTA). Hybridization was carried out overnight at 42°C in the same solution plus 10% w/v dextran sulfate and labelled probe. The blots were rinsed in 2 x SSC, and 0.1% SDS for 5 min, then washed in the same solution for 1 hour at 65°C (1 x SSC is 0.15 M NaCl and 0.015 M sodium citrate). Final washes were in 0.5 x SSC and 0.1% SDS for 1 hour at 65°C. The blots were exposed to Kodak XAR-5 film at -70°C with two Kodak lanex regular intensifying screens.

Subcloning of recombinants was done by cutting the desired fragment out of a low-melting agarose gel and ligating it into appropriately digested DNA from the plasmids pSP64 or pSP65 described in section 2.8 (Melton et al., 1984a; Promega Biotec, Madison, Wisconsin).

2.10 RNA analysis

20 mg of RNA was precipitated in Eppendorf tube, microfuged for 10 min, then the RNA pellet was resuspended in 20 ml of denaturing solution (denaturing solution contained (per 5 ml) 2.5 ml deionized formamide, 0.5 ml 37% formaldehyde, 0.5 ml 10 x MOPS buffer (10 x MOPS buffer is 200 mM MOPS, 50 mM sodium acetate.$3H_2O$, 10 mM EDTA, pH adjusted to 7.0 with acetic acid) and 1.5 ml DEPC H_2O). After heating the sample at 65°C for 15 min the RNA was electrophoresed in 1% (w/v) formaldehyde (0.6%) agarose gel in MOPS buffer in the presence of ethidium bromide (0.25 mg/1 ml). The gel was photographed, and blotted onto Biodyne transfer membrane (PALL) as described by Thomas (1980). After 10 hours the blot was baked at 80°C for 30 min in a vacuum oven then prehybridized and hybridized under the conditions described in section 2.9.

2.11 DNA labelling

Probes were radioactively labelled by nick translation (Rigby et al., 1977) using the Amersham nick-translation kit. In an Eppendorf tube, placed on ice, 1 mg of plasmid DNA was mixed with 20 ml of nucleotide buffer solution (100 mM dATP, 100 mM dGTP, 100 mM dTTP in 250 mM Tris-HCl pH 7.8, 25 mM $MgCl_2$, 50 mM 2-mercaptoethanol), 8 ml [^{32}p] a-dCTP (Amersham, specific activity 3000 Ci/mmole, 1 mCi/ml), 10 ml of enzyme solution (100 pg DNAse I and 5 units E. coli DNA polymerase I in tris-HCl pH 7.5, glycerol, bovine serum albumin and $MgCl_2$) and deionized autoclaved H_2O. The total incubation volume was

100 ml. The tube was gently agitated and placed in a constant temperature water bath at 14.5°C. The progress of the reaction was monitored as follows: 1 ml of the mixture was diluted in 200 ml of H_2O. 20 ml were transferred to a tube containing 50 ml (200 mg) of salmon sperm DNA. After mixing, 3 ml of ice-cold 10% TCA solution was added and the tube was placed at 0°C for 10 min. The precipitated DNA was recovered by vacuum filtration on a Whatman GF/C filter. The filter was washed at least 3 times with 10 ml of 10% TCA solution and dried under vacuum. The filter was counted along with 20 ml of the unprecipitated diluted sample. The probes were labelled typically to a specific activity of 1.5×10^8 cpm/mg DNA. Once this DNA specific activity was achieved the reaction was terminated by adding 90 ml of TE pH 8.0 + 0.1% SDS. Since the percentage of label incorporated was high there was no need to remove the unincorporated nucleotides by G-50 column chromatography. The DNA probes were denatured by boiling for 5 min before use.

CHAPTER III

Cloning and expression of the mouse pgk-1 gene and the nucleotide sequence of its promoter:

RESULTS AND DISCUSSION

3.1 Identifying X-linked pgk-related fragments:

I set out to clone the mouse pgk-1 gene using two pgk-1 cDNA probes kindly supplied to us by Drs. A. Michelson and S. Orkin (Fig. 3.1a). Our earlier attempts to clone the pgk-1 cDNA using oligonucleotide probes had failed (Appendix 3). Probe A is a mouse pgk-1 cDNA containing only part of the pgk-1 mRNA sequence (Appendix 2). The sequencing of this cDNA in collaboration with Dr. P. Boer indicated that it consists of 720 nt corresponding to the 3' end of the pgk-1 mRNA. The sequence is identical to that reported by Mori et al. (1986b). Probe B is a 1.9-kb human pgk-1 cDNA which is nearly the full length of the pgk-1 mRNA (Michelson et al., 1983).

I probed a number of Southern blots of BamHI-digested male and female mouse DNA with the pgk-1 cDNAs (Fig. 3.1b). A number of hybridizing bands are apparent. Most hybridization signals derive from genomic sequences located on autosomal chromosomes because the intensity of most bands is equal in lanes containing DNA isolated from male (even-numbered lanes) and from female (odd-numbered lanes) sources. Since the pgk-1 gene is located on the X chromosome, I concentrated on those bands which appeared twice as intense in female DNA as in male DNA.

Fig. 3.1a: Identification of X-linked pgk-related sequences in the mouse genome. Phosphoglycerate kinase cDNA probes. Probe A is a mouse cDNA containing 720 nt corresponding to the 3' end of the mouse pgk-1 mRNA. Probe B is a human pgk-1 cDNA (Michelson et al., 1983) of almost full length. The filled in areas represent coding regions while the open regions are non-coding. Both cDNA clones were the gifts of Drs. Michelson and Orkin.

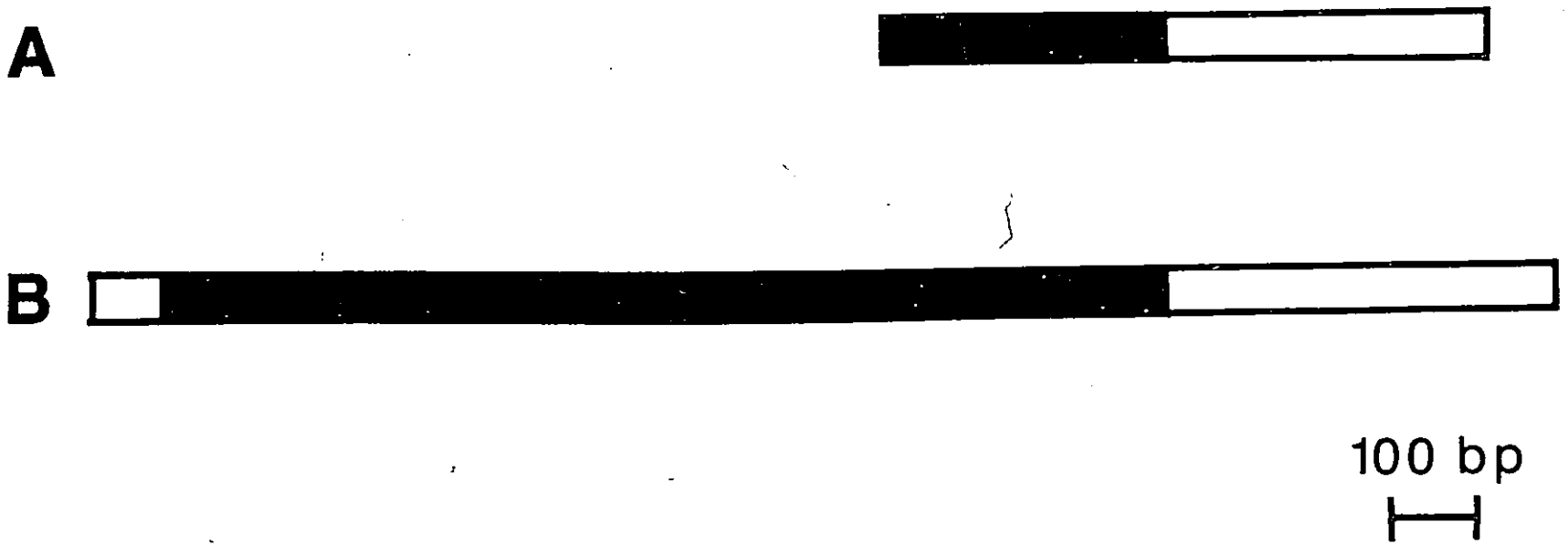
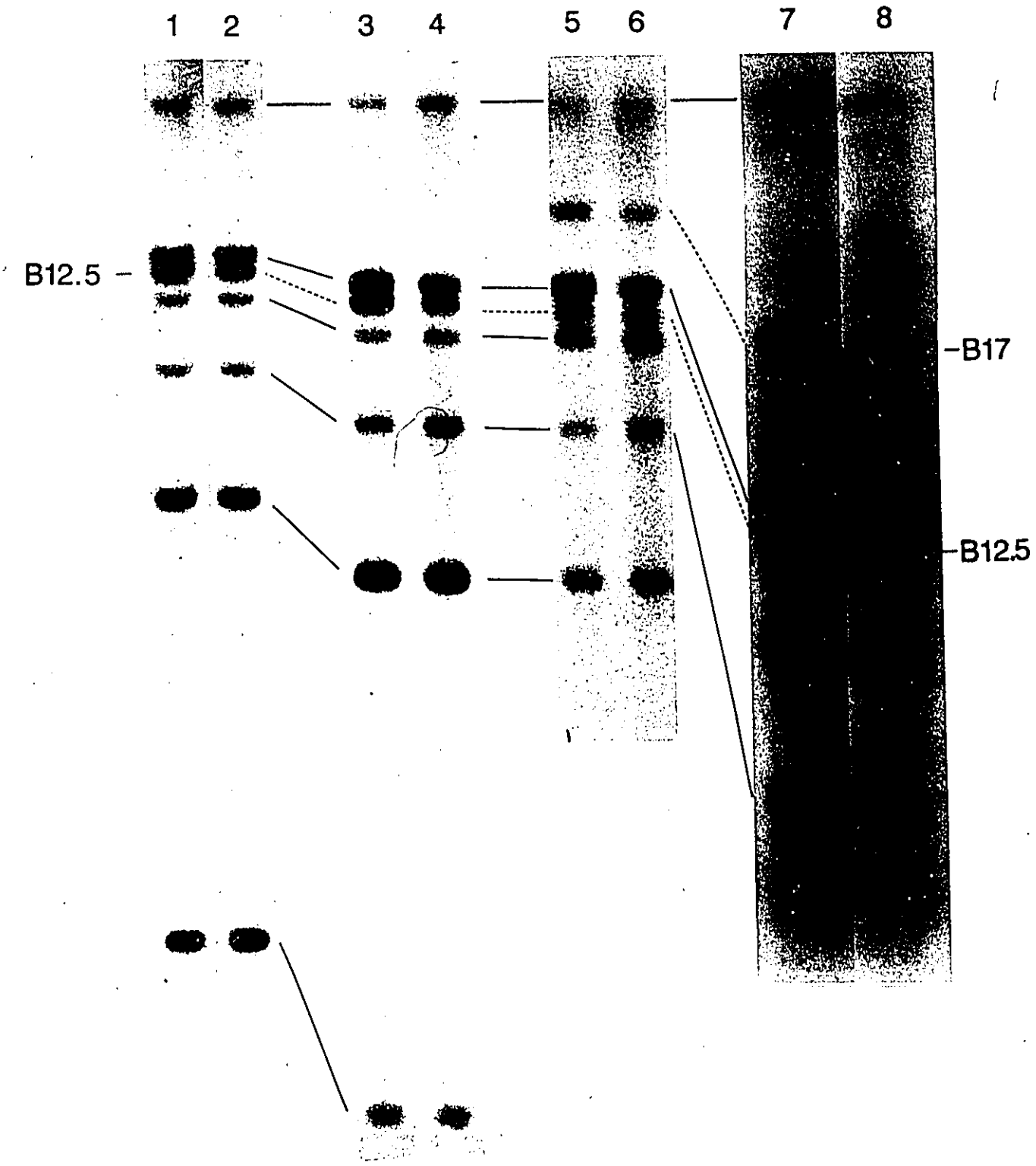

A
B
100 bp

Fig. 3.1b: Southern blot analysis of mouse genomic DNA revealed two X-linked pgk-related fragments. Mouse spleen DNA was digested with BamHI, electrophoresed in 0.6% agarose in 1 x TPE (0.08 M Tris-base, 0.002 M EDTA, 0.08 M phosphoric acid) containing 0.20 mg EtdBr/ml. The DNA was blotted onto Biodyne membranes (PALL, Glen Cove, NY) as described by Maniatis et al. (1982). Prehybridization was carried out in 50% formamide - 5 x SSPE - 2.5 x Denhardt's solution - 0.1% SDS - 0.1 mg/ml denatured salmon testis DNA (1 x SSPE is 0.15 M NaCl, 0.01 M $Na_2H_2PO_4$, 0.001 M EDTA). Hybridization was carried out for 8 h at 42°C in the same solution plus probe A (lanes 1 to 4) or probe B (lanes 5 to 8) which were radioactively labelled by nick translation with ^{32}P-labelled nucleoside triphosphates. All blots were washed at 65°C in 0.1 x SSC - 0.1% SDS (1 x SSC is 0.15 M NaCl - 0.015 M sodium citrate). The blots were exposed to Kodak XAR-5 film at -70°C with two Kodak lanex regular intensifying screens for 24 h. DNA was isolated from female (lanes 1,3,5,7) and male (lanes 2,4,6,8) mice of strain C3H/He (lanes 1 to 6) and Balb/c (lanes 7 and 8). The bands labelled B12.5 and B17 represent BamHI fragments of approximately 12.5 and 17 kb which were consistently more intense in lanes containing female DNA than in corresponding lanes of male DNA.


1
2
3
4
5
6
7
8
B12.5
B17
B12.5

The band labelled B12.5 showed the dosage relationship expected of an X-linked sequence. The B12.5 band was evident in Southern blots hybridized to probe A (Fig. 3.1b, lanes 1 to 4). This B12.5 band was obscured in Southern blots hybridized with probe B due to the appearance of another hybridizing fragment of very similar mobility; however, probe B did reveal an additional band, labelled B17, which was twice as intense in female as in male DNA (Fig. 3.1b, lanes 5 to 8). Because probe A contains pgk-1 sequences derived from the 3' end of the mRNA and probe B is full length, it seemed likely that the B12.5 fragment may contain the 3' end of the pgk-1 gene while the B17 might carry the 5' end.

All of the other hybridizing bands, some of which were polymorphic, were equally intense in both male and female DNA indicating that each of these fragments likely resides on an autosomal chromosome. We have cloned all of these fragments (Chapter 5).

3.2 Cloning the pgk-1 gene:

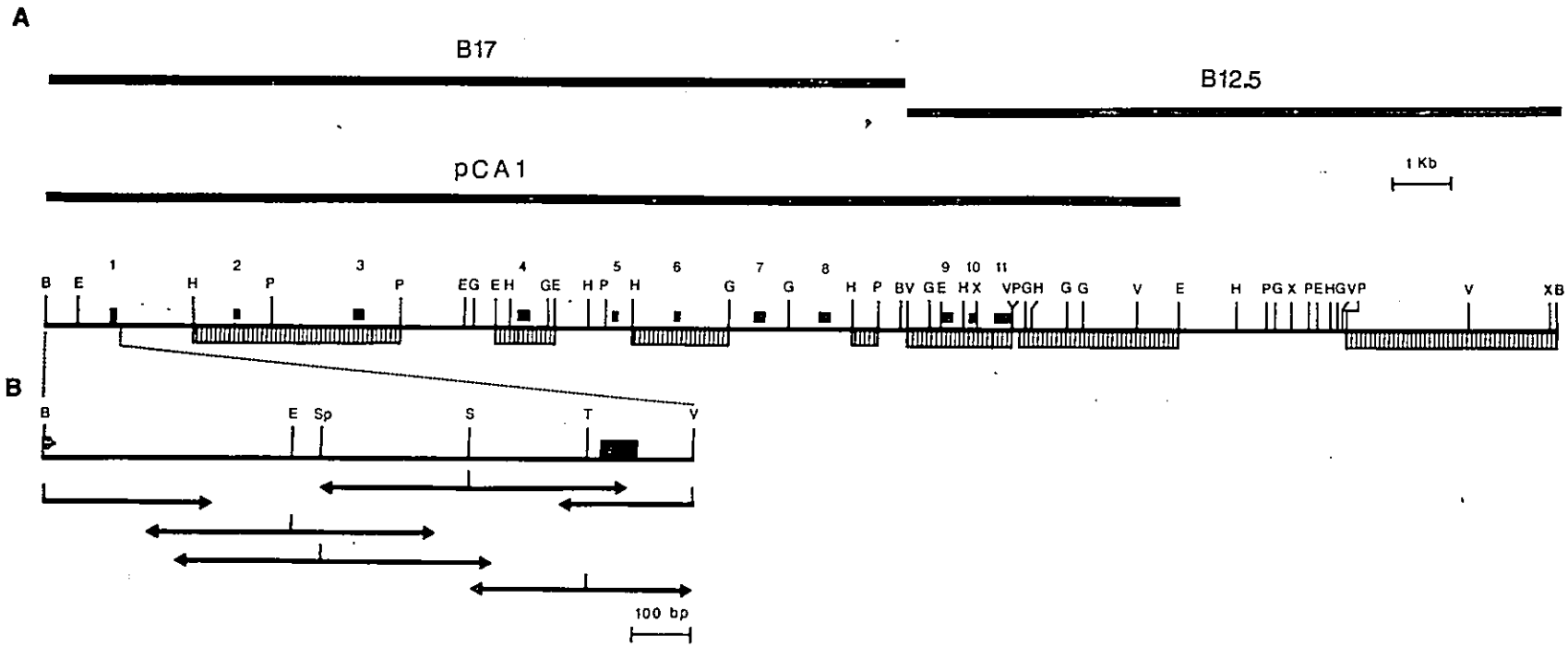
To clone the B12.5 and B17 fragments, I electrophoresed BamHI-digested DNA in agarose gels and the DNA from those regions of the gel containing the B12.5 and B17 fragments were electro-eluted, ligated into the g cloning vector, EMBL3 (Karn et al., 1980; Frischauf et al., 1983) and the libraries were screened with probe B. Cloned recombinant bacteriophages carrying each of the two pgk-1 related sequences were subject to restriction mapping. The digested fragements of the B17 and B12.5 clones were hybridized to probe B to locate regions homologous to the pgk-1 cDNA. The results are shown in

Fig. 3.2A. By making use of the sequence of the mouse pgk-1 cDNA (Mori et al., 1986) and the restriction map, it was possible to deduce that the mouse pgk-1 gene was comprised of at least eleven exons (solid boxes in Fig. 3.2A). These exons are spread over about 16 kb. The exact locations of individual exons within restriction fragments could sometimes be deduced when restriction enzyme recognition sequences were present in the published cDNA sequence (Mori et al., 1986).

To determine whether the B17 and B12.5 fragments were contiguous within the mouse genome, Southern blots of mouse DNA digested with EcoRI or HindIII were hybridized with probes A and B. Fragments spanning the junction between B17 and B12.5 had the expected sizes indicating that the two genomic fragments, B17 and B12.5, are contiguous as shown in Fig. 3.2A.

I hybridized the B17 and B12.5 restriction fragments to radiolabelled total mouse genomic DNA to identify repeated nucleotide sequences. Fragments containing such repeats are indicated with striped boxes in Fig. 2A. The repeated sequences are present within a large number of restriction fragments within and downstream from the pgk-1 gene.

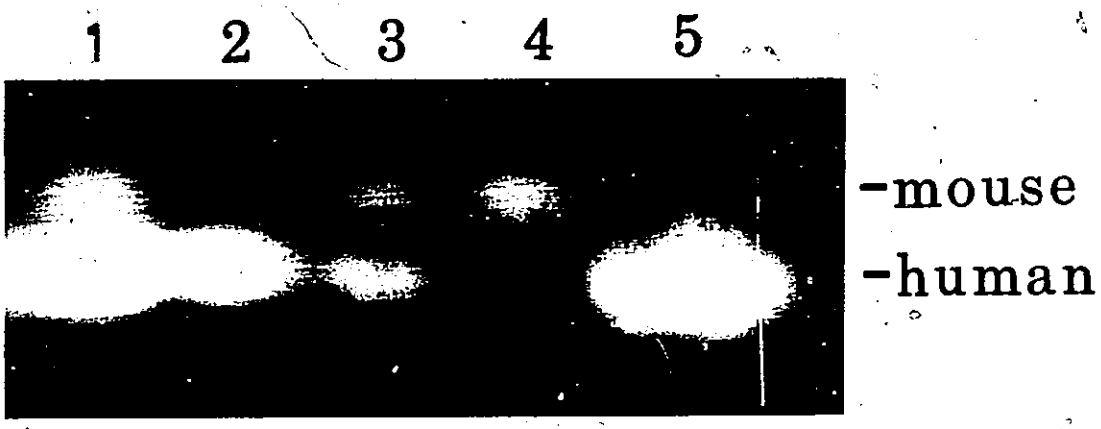
Fig. 3.2: Restriction map of the pgk-1 gene. (Map A) The restriction sites present in the B12.5 and B17 cloned sequences are indicated on the genetic map of the pgk-1 gene along with the extent of DNA present within the B17, B12.5 and pCA1 cloned fragments. The restriction fragments derived from B17 and B12.5 which hybridized to the near full-length cDNA (see Fig. 1a, probe B) are indicated by black boxes while those fragments which contained repeated DNA sequences are indicated with hatched boxes. The restriction sites are: B, Bam HI; E, EcoRI, H, HindIII; P, PstI; G, BglII; X, XbaI; V, PvuII (not all PvuII sites in the B17 fragment were mapped and only the XbaI sites in B12.5 are shown). (Map B) The sequencing strategy for the 5' end of the pgk-1 gene. Fragments were subcloned into plasmids for sequencing by the chain termination method. The arrows indicate the directions and extent of sequence determined from each site. Restriction sites are: Sp, SphI; S, StuI; T, TaqI.


A
B17
B12.5
pCA1
1 Kb
1
2
3
4
5
6
7
8
9
10
11
B E H P P EG EH GE H P H G G H P BV GE HX VPGH GG V E H PG X PEHGVP V XB
B
B E Sp S T V
100 bp

3.3 Expression of PGK-1 from the cloned gene:

To confirm that the B17 and B12.5 fragments comprised the mouse pgk-1 gene, I reconstructed the gene from its cloned fragments and transfected it into human cells in collaboration with Dr. M.W. McBurney to determine if the cloned gene could direct the synthesis of the mouse PGK-1 enzyme. I constructed the pCA1 plasmid (Fig. 3.2A and Appendix 2) from pSP64 (Melton et al., 1984a), it contained the entire pgk-1 gene as well as 0.9 kb of 5'-flanking sequence and 2.8 kb of 3'-flanking sequence. 48 h after transfection of the human cell line, HeLa, with the pCA1 plasmid DNA, a cell extract was prepared, electrophoresed in cellogel strips, and stained for PGK enzyme activity (Bucher et al., 1980; Paterno and McBurney, 1985). The HeLa cell culture transfected with the pCA1 plasmid contained a PGK enzyme activity with mobility similar to that of mouse PGK-1 (Fig. 3.3). HeLa cells transfected with unrelated plasmids contained only the PGK activity with the mobility of the human enzyme. These results confirm that the B17 and B12.5 fragments do indeed carry the mouse pgk-1 gene and that the promoter for this gene is intact and functions following re-introduction into mammalian cells by DNA transfection.

Fig. 3.3: The mouse pgk-1 gene directed expression of mouse PGK-1 activity in transfected human cells. DNA from the pCA1 plasmid (lane 1) or the pSP64 plasmid (lane 2) was transfected into HeLa cells and extracts prepared 48 h later were electrophoresed on cellogel strips and stained for PGK activity (Bucher et al., 1980; Paterno and McBurney, 1985). Lane 3 contained a mixture of mouse and human cell extracts, lane 4 an extract from a strain C3H/He-derived mouse cell line, and lane 5 an extract from HeLa cells. The locations of the mouse and human bands are indicated on the right.


1
2
3
4
5
-mouse
-human

3.4 Cloning the pgk-1^a gene

Nielsen and Chapman, 1977 discovered an electrophoretic variant of PGK-1. This variant existed in Mus musculus in a limited area of Denmark. It is encoded by a variant allele named pgk-1^a. Our studies (Adra et al., 1987; Adra et al., 1988) have indicated that the pgk-1^b allele is associated with the B17 fragment, while a larger fragment, B21, is associated with the pgk-1^a allele (Fig. 3.4). My goal was to clone the pgk-1^a allele to obtain a large sequence flanking the 5' end of pgk-1.

To clone the B21 fragment, I purified DNA from the mouse spleen of a AT10-C3H/HeHa pgk-1^a congenic female kindly supplied by Mr. D. Stephenson and Dr. V. Chapman. BamHI-digested DNA was electrophoresed in agarose gel and the DNA from the region of the gel containing the B21 fragment was electro-eluted. To confirm that the electro-eluted fraction contained the B21 fragment, DNA isolated from the AT10-C3H/HeHa mouse strain was digested with BamHI, electrophoresed in an agarose gel along with a sample from the electro-eluted fraction, and blotted onto a nylon membrane. When probed with a unique fragment subcloned from B17 called pCAB17BH2.5-BE420-64 (Fig. 3.5) the B21 band was evident (Fig. 3.6, lanes 1 and 4). The electro-eluted DNA was ligated into the λ cloning vector, EMBL3 (Karn et al., 1980; Frischauf et al., 1983), packaged and the library was screened with pCAB17BH2.5BE420-64. One reactive plaque was identified. The insert from this hybridizing recombinant lambda genome was characterized based upon fragment size by digesting with Bam HI. Its size was approximately 21 Kbp. The insert was then subcloned into the plasmid Gemini 4 (Fig. 3.7) for restriction mapping and also because

Fig. 3.4: Restriction fragment length polymorphisms of the mouse pgk-1 gene. High-molecular-weight DNA (10 mg) from mouse spleen was digested to completion with BamHI, electrophoresed in 0.6% agarose, and blotted for hybridization. Probe B was used for hybridization to lanes 1-3 containing DNA from a C3H/He female carrying pgk-1^a (lane 1), a female carrying pgk-1^a and pgk-1^b (lane 2) and a female carrying pgk-1^b.

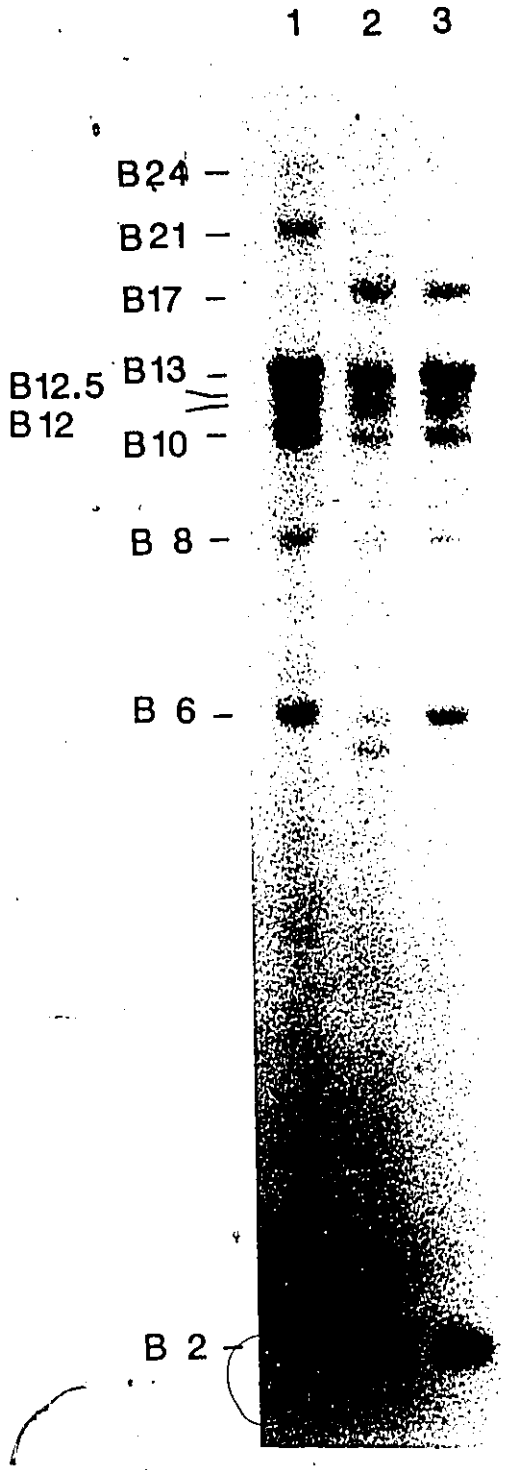
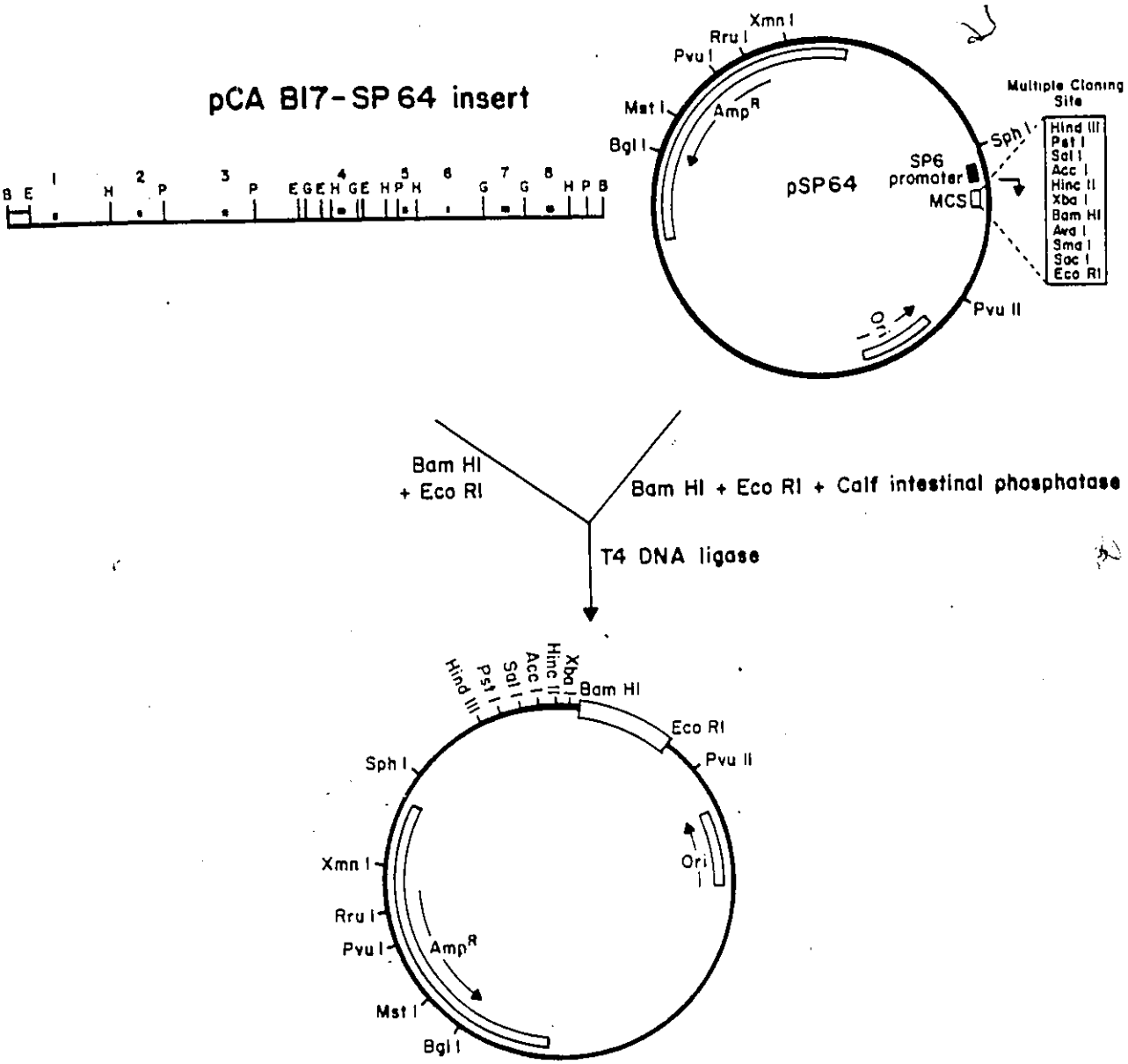

1
2
3
B24
B21
B17
B13
B12.5
B12
B10
B 8
B 6
B 2

Fig. 3.5: pCAB17BH2.5BE420-SP64. This construct contains a unique 420-bpBamHI-EcoRI fragment derived from the 5'-flanking sequence of the mouse pgk-1 gene. This recombinant was constructed by ligating the BamHI/EcoRI fragment (open box) derived from B17 into the equivalent position of pSP64. Only two fragments: B17 and B21, react with this construct when used as a probe (see Fig. 3.6).



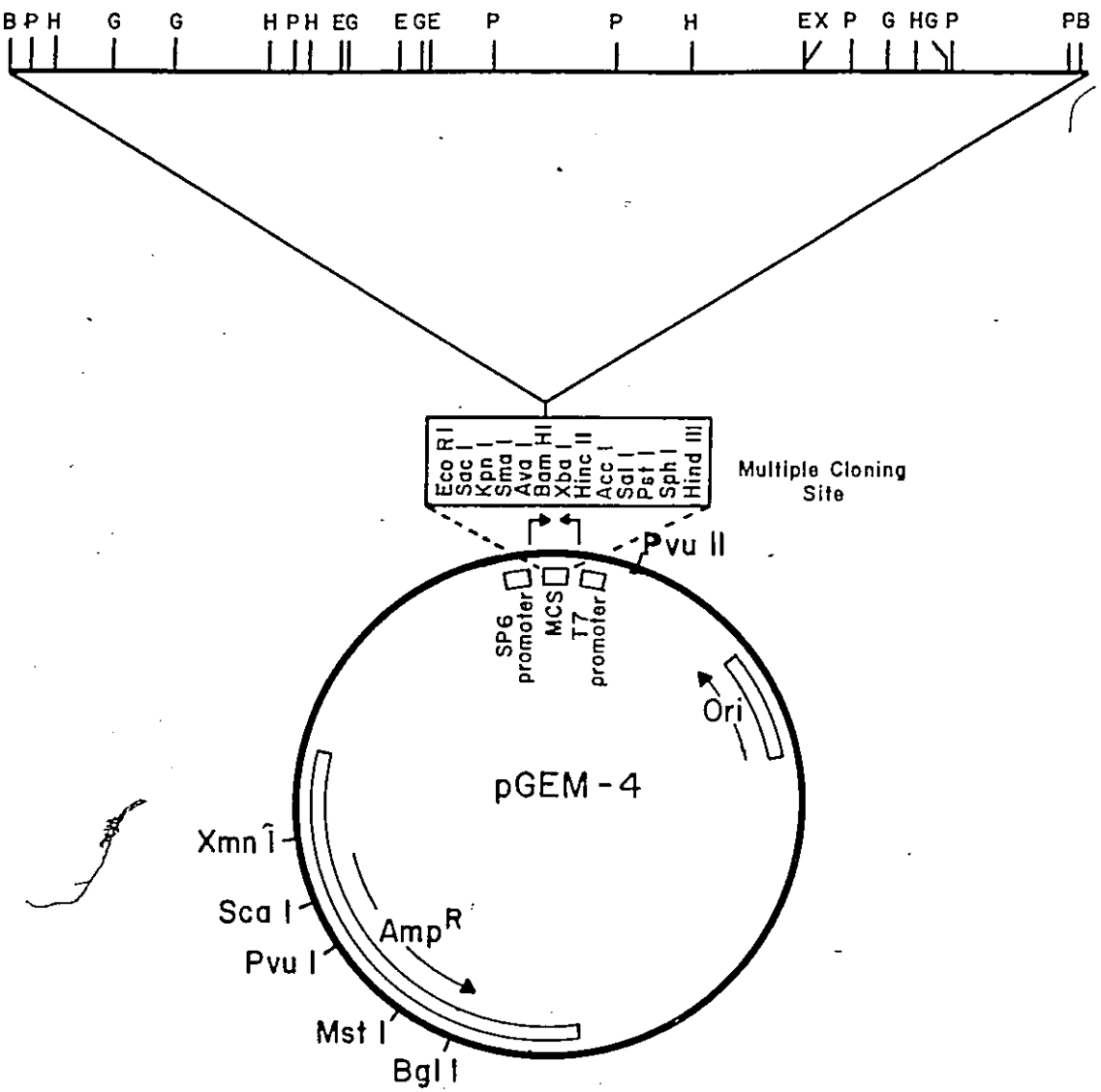
pCA BI7 BH 2.5 BE 420-SP64

Fig. 3.6: Electroelution of B21. To clone the B21 fragment, BamHI-digested DNA, purified from a female carrying pgk-1^a, was electrophoresed in agarose gel and the DNA from the region of the gel containing the B21 fragment was electroeluted. 2 mg of the purified DNA (lane 1) were electrophoresed along with 10 mg of BamHI-digested DNA isolated from P10 XO cells carrying the pgk-1^a allele (lane 2), P10 XX cells carrying pgk-1^a and pgk-1^b (lane 3), AT10-C3H/HeHa pgk-1^a female (lane 4) and C3H/He female carrying pgk-1^b (lane 5).


1 2 3 4 5
B21 –
B17 –

Fig. 3.7: Subcloning of B21 into the plasmid Gemini 4. The genomic library made from the electroeluted DNA containing the fragment B21 was probed with pCAB17BH2.5-BE420-64. One recombinant plaque was identified. The insert from this hybridizing recombinant lambda genome was then subcloned into the BamHI site of the plasmid Gemini 4 using the strategy described in section 2.8.

pCA B2I-GEM-4


B P H G G H P H EG E G E P P H E X P G H G P P B
Eco RI
Sac I
Kpn I
Sma I
Ava I
Bam HI
Xba I
Hinc II
Acc I
Sal I
Pst I
Sph I
Hind III
Multiple Cloning Site
Pvu II
SP6 promoter
MCS
T7 promoter
Ori
pGEM-4
Xmn I
Sca I
Pvu I
Amp R
Mst I
Bgl I

it is more convenient to purify plasmid DNA than phage DNA. I was also concerned about the stability of this large fragment in EMBL3 since its cloning capacity is 9-21 Kbp.

The restriction map of the B17 fragment is identical to that of the B21 fragment within the homologous regions except in three sites (Fig. 3.8, asterisks). The difference in size between the B17 and the B21 fragments results from a BamHI site present in the former fragment and absent in the latter fragment. Also, the second Hind III site in B17 is absent from B21. These two mutations do not affect the amino acid composition of PGK-1 because they are located in non-coding regions. Therefore, they cannot be responsible for the electrophoretic difference between the allelic forms of pgk-1. The third difference affects the 2.25-Kbp Pst I fragment. This fragement is presumed to carry exon 3. It is approximately 2.5 Kbp in B21. This difference in size may have been the result of the movement of a transposon-like element. I intend to investigate the source of this difference and to see if it is responsible for the change in the electrophoretic mobility between the two isozymes.

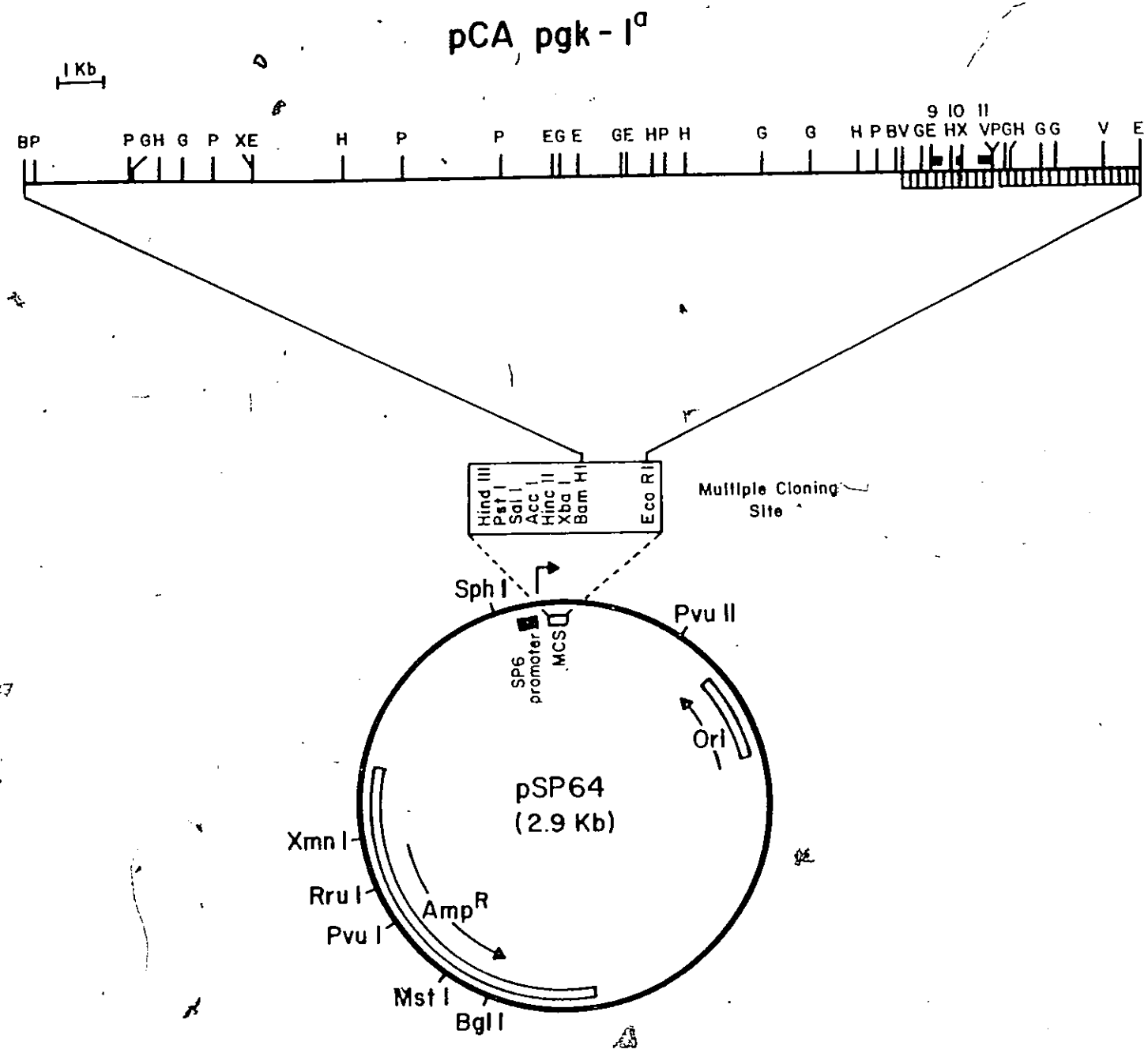
3.5 Expression of PGK-1A from the cloned gene

To determine if amino acid substitution of PGK-1A was due to a codon change in one of first 8 exons, I reconstructed the gene from its cloned fragments. The gene was transfected by Dr. M.W. McBurney into human cells. The pCApgk-1^a plasmid (Fig. 3.9) was constructed from pSP64 (Melton et al., 1984a) and contained the entire B21

Fig. 3.8: Restriction map of the B21 fragment derived from the pgk-1^a allele. The restriction map of B17 (first line) is identical to that of B21 (third line) except in two sites (asterisks). The second line represents the fine map of the 5' end of the pgk-1. The restriction fragments which hybridized to probe B (see Fig. 3.1a) are indicated by black boxes, while those fragments which contained repeated DNA sequences are indicated by striped boxes.

1 Kb
100 bp
1 Kb
B P G H G P X E
H P P P E G E G E H P N G G G H P B
B T T X E Sp S T A
B X E 1 H 2 P J P P E G E H G E H P H 5 6 G 7 G 8 H P B V G E H X 9 10 11 V P G H G G V E H P G X P E H G V P V X B
*
*

Fig. 3.9: Construction of pCApgk-1^a. The recombinant carrying the B12.5 fragments (Appendix 2), called pCA2p-64 was digested to completion with BamHI and partially with EcoRI. A 4.5-kb BamHI/EcoRI fragment containing the last three exons of the pgk-1^b allele was ligated into a BamHI/EcoRI digested pSP64 to generate pCAB12.5 BamHI-EcoRI4.5-64 (Appendix 2). pCAB21-GEM-4 (Fig. 3.7) was digested with BamHI, then the B21 fragment was ligated to pCAB12.5 BamHI-EcoRI4.5-64 cleaved with BamHI and treated with calf intestinal phosphatase. Orientation of the DNA insert was determined by restriction analysis (data not shown).


pCA pgk - I^a
1 Kb
BP
P GH G P XE
H
P
P
EG E
GE HP H
G
G
H P BV GE HX VPGH GG
9 10 11
V
E
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Bam HI
Eco RI
Multiple Cloning Site
Sph I
Pvu II
SP6 promoter
MCS
Ori
pSP64
(2.9 Kb)
Xmn I
Rru I
Pvu I
Amp^R
Mst I
Bgl I

fragment plus a 4.5 Kb fragment containing exons 9, 10 and 11 of the pgk-1^b allele. This construct directed the expression of mouse PGK-1A activity (data not shown).

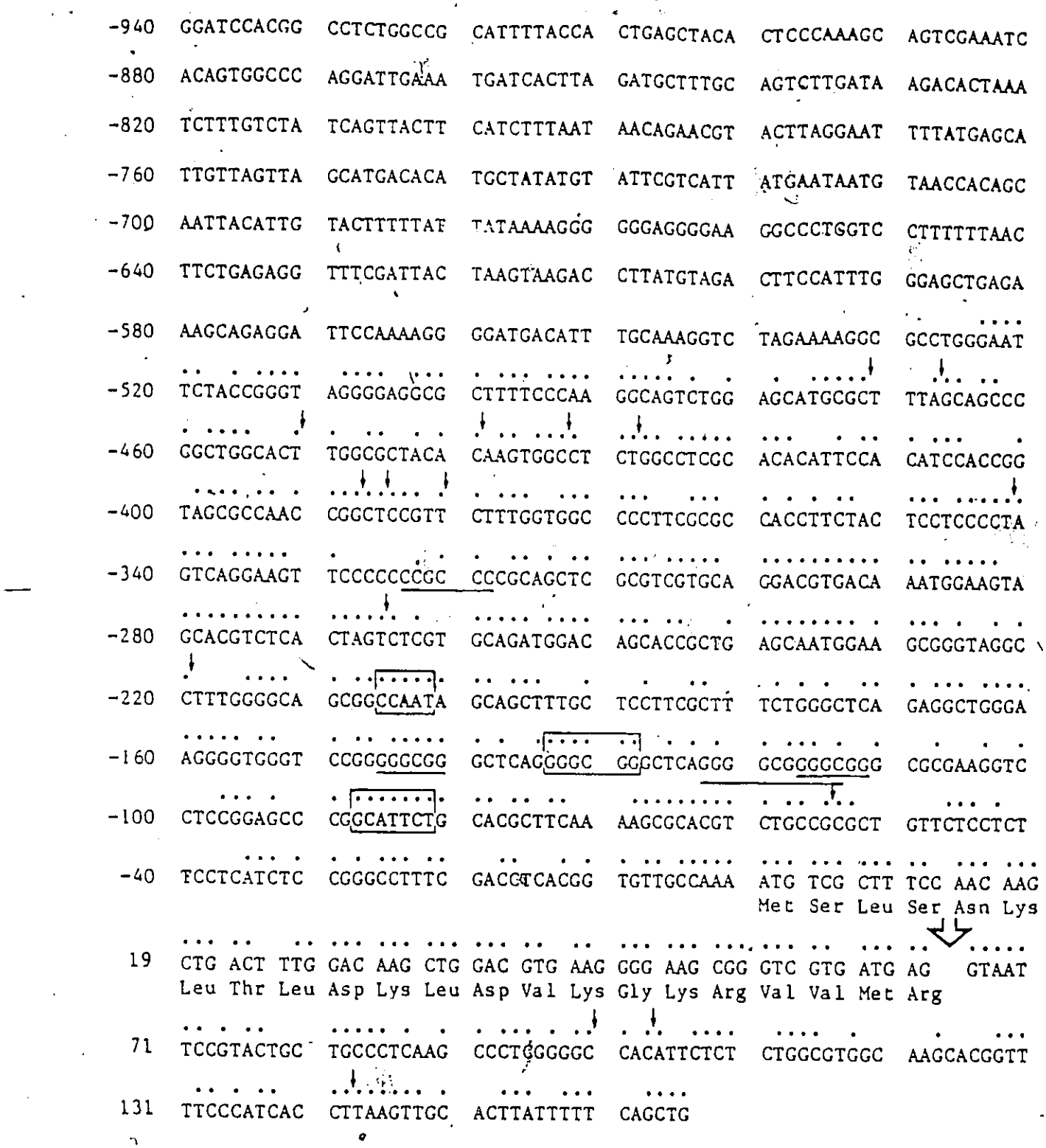
3.6 The pgk-1 promoter

The 5'-flanking region and first exon of the mouse pgk-1 gene were sequenced in collaboration with Dr. P. Boer according to the strategy outlined in Fig. 3.2b. The sequence is shown in Fig. 3.10. Nucleotides are numbered from the translation start site.

The sequence from nt -21 to +65 is identical (except for one mismatch at nt -9) to the mouse pgk-1 cDNA sequence of Mori et al. (1986). Homology with the cDNA ends after nt +65 indicating that this is the 5' end of the first intron. The deduced amino acid sequence of this first exon and the position of the first intron in the mouse pgk-1 gene are identical to those of the human pgk-1 gene (Singer-Sam et al., 1984).

The sequence of the mouse pgk-1 5'-flanking region contains sequence motifs similar to those found in the promoters of a variety of other constitutively expressed genes (Melton et al., 1984b; Patel et al., 1986; Martini et al., 1986; Broderick et al., 1987). There are five GC boxes (Dynan and Tjian, 1985) which are the presumptive binding sites for the Sp1 transcription factor. One of the GC boxes, GGGGCGGGGC at nt -123, has the preferred sequence for highest affinity Sp1 binding (Kadonaga et al., 1986). A CCAAT motif is present at nt -206 and a putative transcription start point is present at nt -88 (Corden et al., 1980). There is no TATA box.

Fig. 3.10: Nucleotide sequence of the 5' end of the mouse pgk-1 gene. The sequence was obtained by the strategy outlined in Fig. 3.2B. The sequence is numbered from the translation start site and the amino acid sequence of the first exon is indicated. The large open downward arrow indicates the first intron donor splice site. Underlined sequences at nt -324, nt -146, nt -123, and nt -117 represent possible Sp1 binding sites. The boxed sequences are the presumed functional motifs whose sequences are conserved between human and mouse pgk-1 genes: a CCAAT box at nt -206, a GC box at nt -134, and a putative transcription start point at nt -88. The human pgk-1 upstream sequence was compared to the mouse sequence starting from the EcoRI site present in both sequences at nt -524 through the first exon to nt +166. Identical nucleotides in the human pgk-1 sequence are indicated by dots above the mouse sequence and inserted groups of 1 to 7 nt in the human sequences are indicated by small arrows.


-940 GGATCCACGG CCTCTGGCCG CATTTTACCA CTGAGCTACA CTCCCAAAGC AGTCGAAATC
-880 ACAGTGGCCC AGGATTGAAA TGATCACTTA GATGCTTTGC AGTCTTGATA AGACACTAAA
-820 TCTTTGTCTA TCAGTTACTT CATCTTTAAT AACAGAACGT ACTTAGGAAT TTTATGAGCA
-760 TTGTTAGTTA GCATGACACA TGCTATATGT ATTCGTCATT ATGAATAATG TAACCACAGC
-700 AATTACATTG TACTTTTTAT TATAAAAGGG GGGAGGGGAA GGCCCTGGTC CTTTTTTAAC
-640 TTCTGAGAGG TTTCGATTAC TAAGTAAGAC CTTATGTAGA CTTCCATTTG GGAGCTGAGA
-580 AAGCAGAGGA TTCCAAAAGG GGATGACATT TGCAAAGGTC TAGAAAAGGC GCCTGGGAAT
-520 TCTACCGGGT AGGGGAGGCG CTTTTCCCAA GGCAGTCTGG AGCATGCGCT TTAGCAGCCC
-460 GGCTGGCACT TGGCGCTACA CAAGTGGCCT CTGGCCTCGC ACACATTCCA CATCCACCGG
-400 TAGCGCCAAC CGGCTCCGTT CTTTGGTGGC CCCTTCGCGC CACCTTCTAC TCCTCCCCTA
-340 GTCAGGAAGT TCCCCCCCGC CCCGCAGCTC GCGTCGTGCA GGACGTGACA AATGGAAGTA
-280 GCACGTCTCA CTAGTCTCGT GCAGATGGAC AGCACCGCTG AGCAATGGAA GCGGGTAGGC
-220 CTTTGGGGCA GCGGCCAATA GCAGCTTTGC TCCTTCGCTT TCTGGGCTCA GAGGCTGGGA
-160 AGGGGTGGGT CCGGGGGCGG GCTCAGGGGC GGGCTCAGGG GCGGGGCGGG CGCGAAGGTC
-100 CTCCGGAGCC CGGCATTCTG CACGCTTCAA AAGCGCACGT CTGCCGCGCT GTTCTCCTCT
-40 TCCTCATCTC CGGGCCTTTC GACGTCACGG TGTTGCCAAA ATG TCG CTT TCC AAC AAG
Met Ser Leu Ser Asn Lys
19 CTG ACT TTG GAC AAG CTG GAC GTG AAG GGG AAG CGG GTC GTG ATG AG GTAAT
Leu Thr Leu Asp Lys Leu Asp Val Lys Gly Lys Arg Val Val Met Arg
71 TCCGTACTGC TGCCCTCAAG CCCTGGGGGC CACATTCTCT CTGGCGTGGC AAGCACGGTT
131 TTCCCATCAC CTTAAGTTGC ACTTATTTTT CAGCTG

The region containing the GC boxes has an unusual sequence structure. The 12-nt motif, GGGGCGGGCCTCA, is present in tandem copies at nt -147 and -135 and is followed at nt -123 by a third repeat of the first 8 nt of this motif.

The region between nt -540 and nt +130 is very rich in the dinucleotide CpG (Bird et al., 1985). There are 50 CpGs within this 670-nt region but only six in the 400 nt from -940 to -540.

The mouse pgk-1 promoter sequence was compared to that of the analogous human gene (Singer-Sam et al., 1984). When compared over the region available (-524 to +166), and allowing for looping out of 1 to 7 nt (Queen and Korn, 1983) the sequences showed 68% nt identity (identical nucleotides are indicated as dots in Fig. 4 and sites of extra bases in the human sequence are shown as arrows). The coding region was 91% homologous. The presumed functional motifs, the CCAAT box and putative transcription start point, were particularly homologous, whereas only the one GC box at -144 appeared to be conserved. The spacing between the homologous regions was also well conserved. Several other regions of high sequence conservation were also noted; in particular, the 47-nt region from -312 to -266 shows 90% nucleotide identity. Regions of such high nucleotide identity strongly suggest some functional significance which may be revealed by expression studies; however, this 47-nt motif is not present in the mouse hprt promoter (Melton et al., 1984b) so it does not appear to be a characteristic of X-linked genes.

In preliminary analysis of this promoter region, I subcloned the sequence spanning -523 to -22 into a plasmid in front of the *E. coli* lacZ gene (Hall et al., 1983). Following transfection into mouse

fibroblasts and embryonal carcinoma cells, this plasmid directed the synthesis of β-galactosidase activity (R. Berg, C.N.A., P.H.B. and M.W. McB., unpublished observations). Thus, the promoter sequence reported in Fig. 4 functions to direct the transcription of a heterologous gene.

The yeast pgk promoter is one of the strongest in that organism (Ogden et al., 1986; Chen and Hitzeman, 1987) and our preliminary experiments suggest that the mouse pgk-1 promoter also has strong promoter activity when transfected into mammalian cells. In addition, pgk-1 mRNA and enzymatic activity are present in all somatic tissues suggesting that the pgk-1 promoter is active in most cell types. Thus, this promoter may be a valuable one for ensuring constitutive expression in cells transfected or microinjected with genes whose expression one wishes to maintain relatively constant in all cell types.

3.7 Conclusions

The structure of many genes is conserved in different mammalian species and the pgk-1 gene appears to be no exception. Michelson et al. (1985) showed that the human pgk-1 gene consists of eleven exons contained within a 23-kb region of the human X chromosome. The mouse pgk-1 gene is spread over 16 kb and contains eleven or more exons. The first exon in both human and mouse pgk-1 genes encodes an identical sequence of amino acids. Hybridization of subcloned fragments of the cDNA probes to restriction fragments of B17 and B12.5 (Adra et al., 1988) are consistent with the idea that the number and

positions of all exons of the pgk-1 gene are the same in mouse and human.

The major features of the mouse pgk-1 promoter are similar to those of a variety of other X-linked and autosomally encoded genes whose expression is not restricted to particular tissue types. Like pgk-1, genes such as hprt (Melton et al., 1984b; Patel et al., 1986), g6pd (Martini et al., 1986), aprt (Broderick et al., 1987), ada (Ingolia et al., 1986), and tk (Flemington et al., 1987) all have multiple GC boxes within their promoters, they lack TATA boxes, and may or may not have a CAAT box. All of these promoters contain a high proportion of CpG dinucleotides which are potential sites for DNA methylation.

Studies comparing the extent of DNA methylation in the CpG-rich regions of a number of X-linked genes have indicated that these sites on the inactive X chromosome DNA are heavily methylated while the same sites on DNA from the active X chromosome are methylated to a much lesser degree (Wolf et al., 1984; Yen et al., 1984; Lindsay et al., 1985; Lock et al., 1986; Keith et al., 1986). However, DNA methylation is unlikely to be solely responsible for inactivation because (i) genes expressed in a tissue-specific fashion do not have CpG islands, there are no apparent methylation differences between such genes on the active and inactive X chromosomes (Cullen et al., 1986) and they are not reactivated by treatment with 5-AC (see Appendix 1). (ii) the inactive X chromosome in normal human diploid cells (Wolf and Migeon, 1982) and in some embryonal carcinoma cell lines (Paterno et al., 1985) cannot be reactivated with DNA demethylating agents, and (iii) the kinetics of X inactivation and

methylation are different in differentiating embryonic and embryonal carcinoma cells (Lock et al., 1987). Thus DNA methylation, perhaps in concert with other mechanisms, may act to maintain the inactive state but some other event(s) is responsible for its initiation.

Clues to the mechanism by which X inactivation is initiated may come from analysis of the X-linked "controlling element" locus in the mouse called Xce. Different alleles of Xce act in a cis-dominant fashion to determine the probability that an X chromosome remains active in each female somatic cell (Cattanach, 1975; Johnson and Cattanach, 1981; Rastan, 1982). Although no direct means of identifying the Xce locus is available, the pgk-1 gene is very closely linked to this locus (Cattanach and Papworth, 1981) and it may be possible to use chromosome walking techniques to identify this region.

CHAPTER IV

The Role of DNA Methylation in X Chromosome Inactivation

4.1 Introduction

The DNA of most eukaryotic cells is subject to one postreplicational modification. This is the methylation of the cytosine in the dinucleotide CpG (reviewed by Riggs and Jones, 1983; Razin and Szyf, 1984; Jaenisch and Jähner, 1984; Bird, 1986). Several models have been proposed suggesting a role for DNA methylation in X chromosome inactivation (Riggs, 1975; Holliday and Pugh, 1975; Sager and Kitchin, 1975). These models are supported by three lines of evidence. First, treatment of cells with DNA-demethylating agents can result in the expression of genes on the inactive X chromosome (see Section 1.15.1.1; Mohandas et al., 1981; Lester et al., 1982; Graves, 1982; Hors-Cayla et al., 1983; Paterno et al., 1985; Ellis et al., 1987). Second, DNA-mediated gene transfer experiments showed that the hprt gene on the inactive X chromosome does not transform hprt$^-$ recipient cells unless it had been previously reactivated by 5-azacytidine (Liskay and Evans, 1980; Chapman et al., 1982; Lester et al., 1982; Venolia and Gartler, 1983; also see Section 1.15.1.2). Third, the CpG-rich regions of several X-linked housekeeping genes are hypermethylated in the transcriptionally inactive state (Yen et al., 1984; Wolf et al., 1984; Toniolo et al., 1984; Lindsay et al., 1985; Lock et al., 1986; Keith et al., 1986; also see Section 1.15.1.3).

To examine the relationship between gene inactivity and

methylation we investigated the methylation pattern at HpaII and HhaI sites present in the 5' flanking region of the mouse pgk-1 gene in kidney, spleen and EC cell lines representing different stages of X chromosome differentiation (Paterno et al., 1985). We chose the pgk-1 gene because it is subject to dosage compensation. In addition, it is expressed in undifferentiated EC cells and in all somatic cells.

Our analysis of the upstream region indicated that the CCGG sites assayed for methylation are unmethylated on both the active and inactive X chromosomes.

Results

4.2 The methylation state of the 5' pgk-1 GpG rich island in adult mice

The 5'-flanking region of the pgk-1 gene contains a number of potential methylation sites within the recognition sequences for the restriction endonucleases MspI and HpaII (Fig. 4.1). Both enzymes recognize the same sequence: 5'-CCGG-3'. Whereas HpaII cuts this sequence only in the absence of a methyl group on the internal cytidine, its isoschizomer MspI is insensitive to the methylation status of this internal C (McClelland, 1981). Since the 5' regions of many inactive genes are extensively methylated we examined the methylation status of the upstream region of mouse pgk-1 in spleen and kidney DNA from 26 days old female and male mice. To examine the methylation level I used the restriction endonucleases HpaII or MspI in combination with BamHI and PvuII (Fig. 4.2). When the BamHI/PvuII digested DNA was probed, a 1106 bp band was detected. This band was

Fig. 4.1: Restriction map of the 1106bp BamHI-PvuII fragment containing the 5' end of the mouse pgk-1 gene. The first exon is indicated by a black box. Restriction sites are: B, BamHI; PV, PvuII; , HhaI sites; , HpaII sites. The numbers above and below the line indicate the size of the expected HhaI and HpaII fragments respectively.

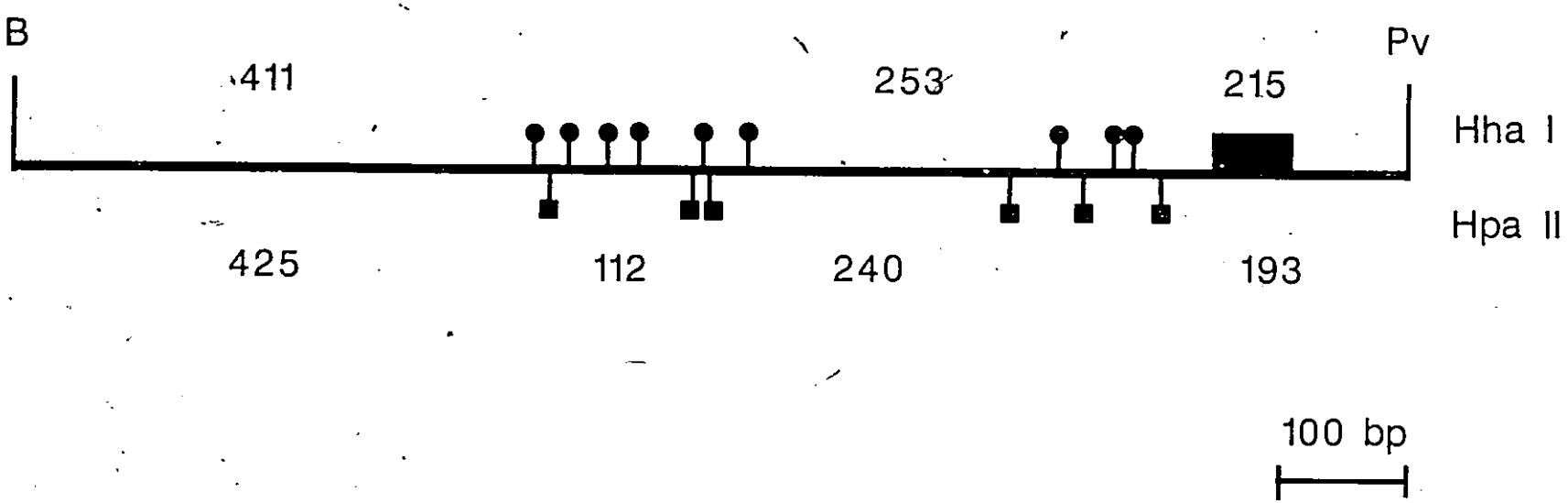
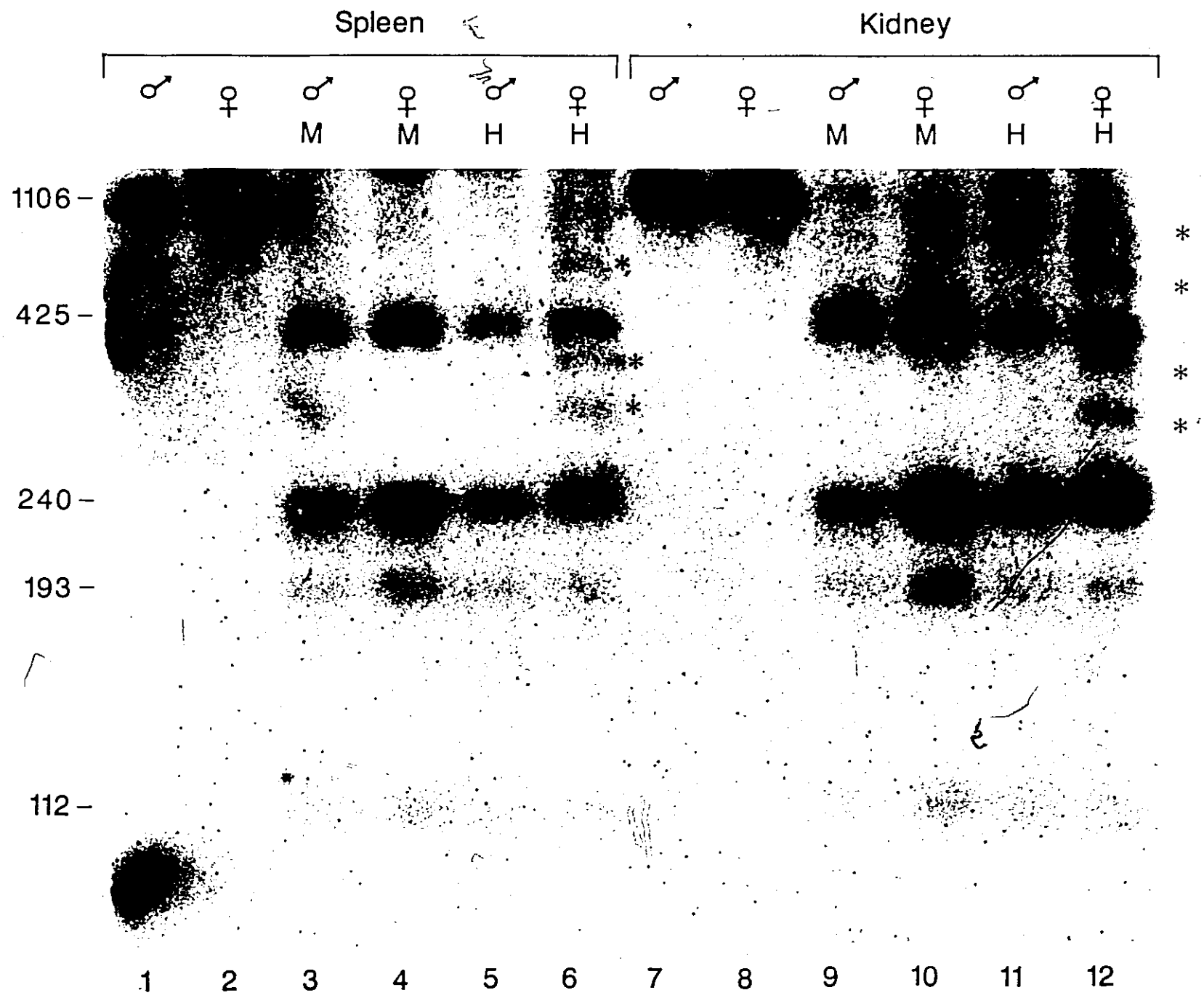

B
411
253
215
Pv
Hha I
Hpa II
425
112
240
193
100 bp

Fig. 4.2: Southern analysis showing lack of significant differential methylation of CCGG sites of the 5' end of the pgk-1 gene. Spleen and kidney DNA (15 mg) was digested with BamHI + PvuII (lanes 1, 2, 7 and 8) or with BamHI + PvuII + MsPI (Lanes 3, 4, 9 and 10) or with BamHI + PvuII + HpaII (lanes 5, 6, 11 and 12). The digestion was carried out for 4-20 hours using at least fivefold excess of enzyme under conditions recommended by the manufacturer. Completeness of the restriction enzyme digestions was tested and confirmed by mixing pSP64 DNA with an aliquot of each reaction mixture. The digested DNA was electrophoresed in a 4% NuSieve agarose (FMC, Rockland, ME) / 0.2% BRL agarose (Bethesda Research Laboratories) gel in 1XTAE buffer containing 0.20 mg EtdBr/ml. The DNA was blotted onto Biodyne membranes as described by Maniatis et al. (1982) and crosslinked by UV light. Prehybridization was carried out in 50% formamide-5 x SSPE -2.5 x Denhardt's solution - 0.1% SDS - 0.2 mg/ml denatured salmon testis DNA (1 x SSPE is 0.15 M NaCl, 0.01 M $Na_2H_2PO_4$, 0.001 M EDTA). Hybridization was carried out for 8 h at 42°C in the same solution plus two probes, covering the entire BamHI-PvuII fragment, which were radioactively labelled by nick translation with ^{32}P-labelled nucleoside triphosphates. The blot was washed at 65°C in 0.1 x SSC - 0.1% SDS (1 x SSC is 0.15 M NaCl - 0.015 M sodium citrate). The blot was exposed to Kodak XAR-5 film at -70°C with two Kodak lanex regular intensifying screens for 24 h.


Spleen
Kidney
♂
♀
♂ M
♀ M
♂ H
♀ H
♂
♀
♂ M
♀ M
♂ H
♀ H
1106 –
425 –
240 –
193 –
112 –
*
*
*
*
*
*
1
2
3
4
5
6
7
8
9
10
11
12

twice as intense in female as in male DNA (Fig. 4.2, lanes 1, 2, 7 and 8). BamHI/PvuII/ MspI digested DNA from both sexes exhibited four bands of 112, 193, 240 and 425 bp (Fig. 4.2, lanes 3, 4, 9 and 10). The remaining three fragments (60, 59 and 14 bp) could not be detected because they are very small. In search of differences between the male and female somatic DNA the HpaII/BamHI/PvuII digested DNA was probed. Clearly, the pgk-1 gene on the inactive X chromosome is not extensively methylated, as DNA digests (lanes 5, 6, 11 and 12) contained the same fragments in males and females, a 1106 bp band was not seen, and fragments were more intense in females than males. After prolonged exposure four new bands were detected in female DNA (lanes 6 and 12). The size of the top band indicates that all HpaII sites, except site 6 (sites are numbered beginning with the most 5' site) which is located in exon 1, can be partially methylated on the inactive X chromosome.

The possibility that these results could have risen from a partial HpaII digestion is very unlikely. The digests were monitored with an internal control of marker DNA which was digested to completion in all cases. Furthermore, the experiment was repeated several times with excess enzymes and the same digestion pattern was seen. The data showed that the entire 1106 bp fragment of the 5' end of the mouse pgk-1 gene on the inactive X chromosome was not significantly methylated in adult female mice. However both cytogenetic and biochemical data indicated that the X inactivation process is completed in the embryonic lineages by day 6.5 in mouse embryogenesis (Monk and Harper, 1979; Takagi et al., 1982; also see Section 1.3).

4.3 Methylation of the 5' pgk-1 CpG rich island in various EC cells

Both X chromosomes are genetically active in early female embryos (Takagi, 1974; Kratzer and Gartler 1978; Monk and Harper, 1978; also see Section 1.2). The differentiation of the embryonic cells is accompanied by the inactivation of one X chromosome in each cell (reviewed by Monk, 1978; Gartler and Riggs, 1983). Female mouse embryonal carcinoma cells can be used to study this process of X chromosome inactivation because certain EC cells, such as P10 and LT-1, carry two active X chromosomes. The differentiation of these cells in culture is accompanied by the inactivation of one X chromosome in each cell (Martin et al., 1978; McBurney and Strutt, 1980; Takagi and Martin, 1985; Paterno and McBurney, 1985; also see Sections 1.10.1.1 and 1.10.1.3). Other female EC cells appear to be frozen at different stages of X chromosome differentiation (Paterno et al., 1985). AGM2, C100, and C145 cell lines carry an inactive X chromosome. Cytogenetic studies demonstrated the presence of an allocyclic X chromosome in AGM2 and C145 lines. However, allocyclic replication of X chromosomes was not present in C100 EC cell line indicating that it may be arrested at an intermediate stage of X chromosome differentiation. The status of the X chromosomes in these female EC cell lines and the fact that C145 cells, unlike C100 and AGM2 cells, do not reactivate the inactive X following exposure to 5-azacytidine support the view that these three cell lines from C100 to C145 represent different stages of X chromosome inactivation. The lack of correlation between methylation and gene expression in the 5' region of pgk-1 does not exclude the possibility that methylation may

be a likely mechanism which marks a single X chromosome to be inactivated by some other regulatory mechanisms. Following this event methylation may no longer be required.

To detect a possible shift from the methylated to the undermethylated state during embryogenesis I investigated the extent of methylation of the 5' control region of the mouse pgk-1 gene in the various EC cell lines described in Chapter I. Blot analysis of DNA cleaved with HhaI showed that sites 1, 7 and 9 (sites are numbered beginning with the most 5' site) (see Fig. 4.1) are unmethylated in AGM2 DNA (Fig. 4.3, lane 3), C100 DNA (lane 4), C145 DNA (lane 5) and even in female somatic cells with Xi (lane 2). The fragments produced were similar to those in males (lane 1) but their intensity was different. The clustering of site 8 to 7 and 9 and of sites 2, 3, 4 and 5 to one another made the examination of their methylation state almost impossible. However, the methylation of sites 2, 3, 4 and 5 throughout the cluster would have yielded a detectable fragment.

High molecular weight genomic DNA was digested with HpaII or MspI in combination with BamHI and PvuII. Southern blot hybridization of the digests demonstrated that the HpaII banding patterns were essentially identical to those generated by MspI (Fig. 4.4). Therefore, the DNA in the reactivated clone AGM2 reac-2a (Fig. 4.4, lanes D) is like the DNA in AGM2 (lanes A), C100 (lanes B), C145 (data not shown) and differentiated AGM2 cells (lanes C), being fully unmethylated at the HpaII sites in the 5'-flanking sequence of pgk-1. Not all of the expected HpaII fragments (60, 59 and 14 bp) were visible in the autoradiographs. However, we can conclude that these sites are not methylated because fragments larger than the ones predicted from the nucleotide sequence would have been observed.

Fig. 4.3: Southern blot analysis of BamHI + PvuII + HhaI digestion of various DNAs indicating the lack of female-specific restriction fragments in the 5' end of pgk-1. Genomic DNA (15 mg) was digested to completion, blotted and hybridized as described previously in the legend of Fig. 4.2. 1, male mouse kidney DNA; 2, female mouse kidney DNA; 3, AGM2 DNA; 4, C100 DNA; 5, C145 DNA.

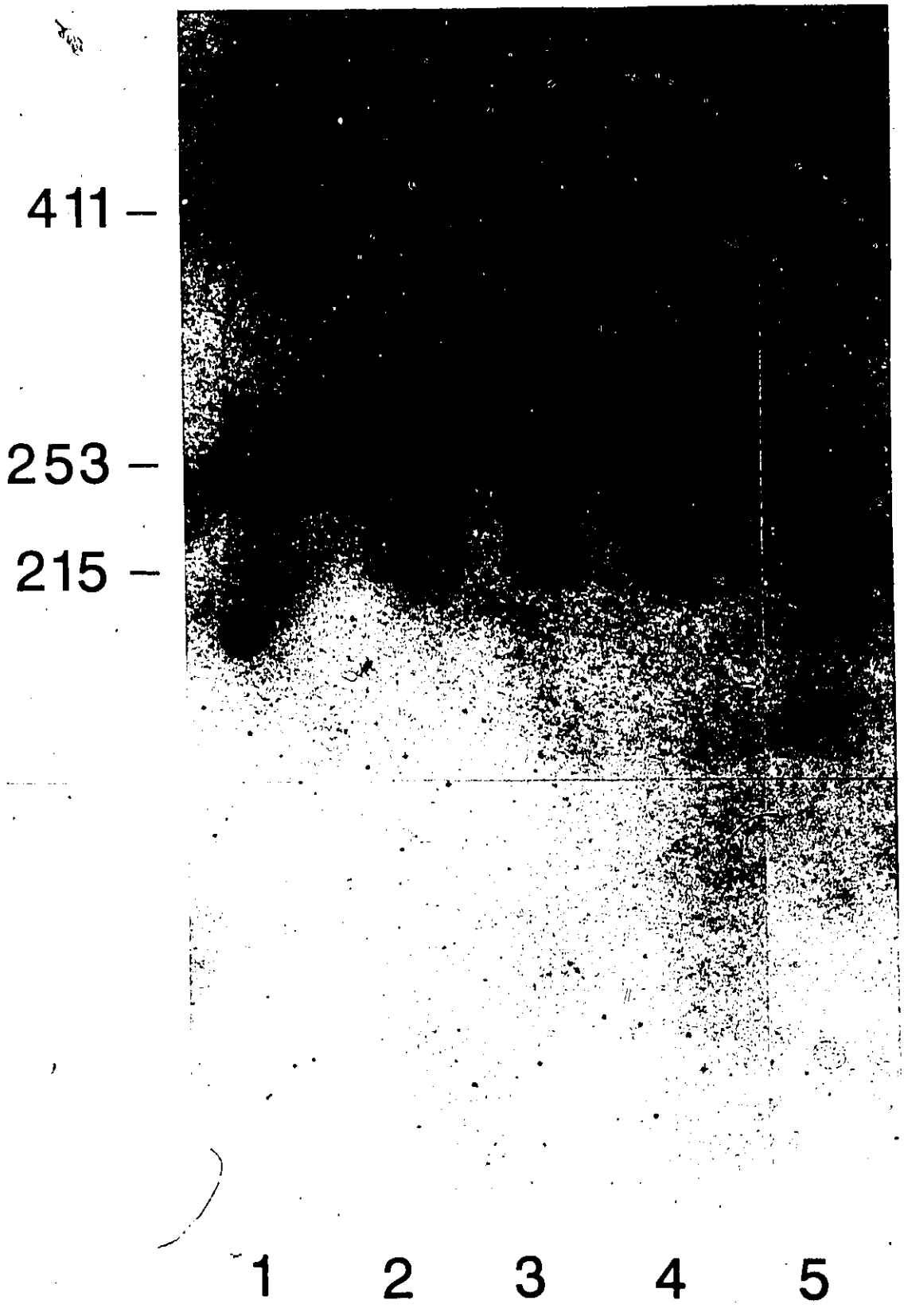
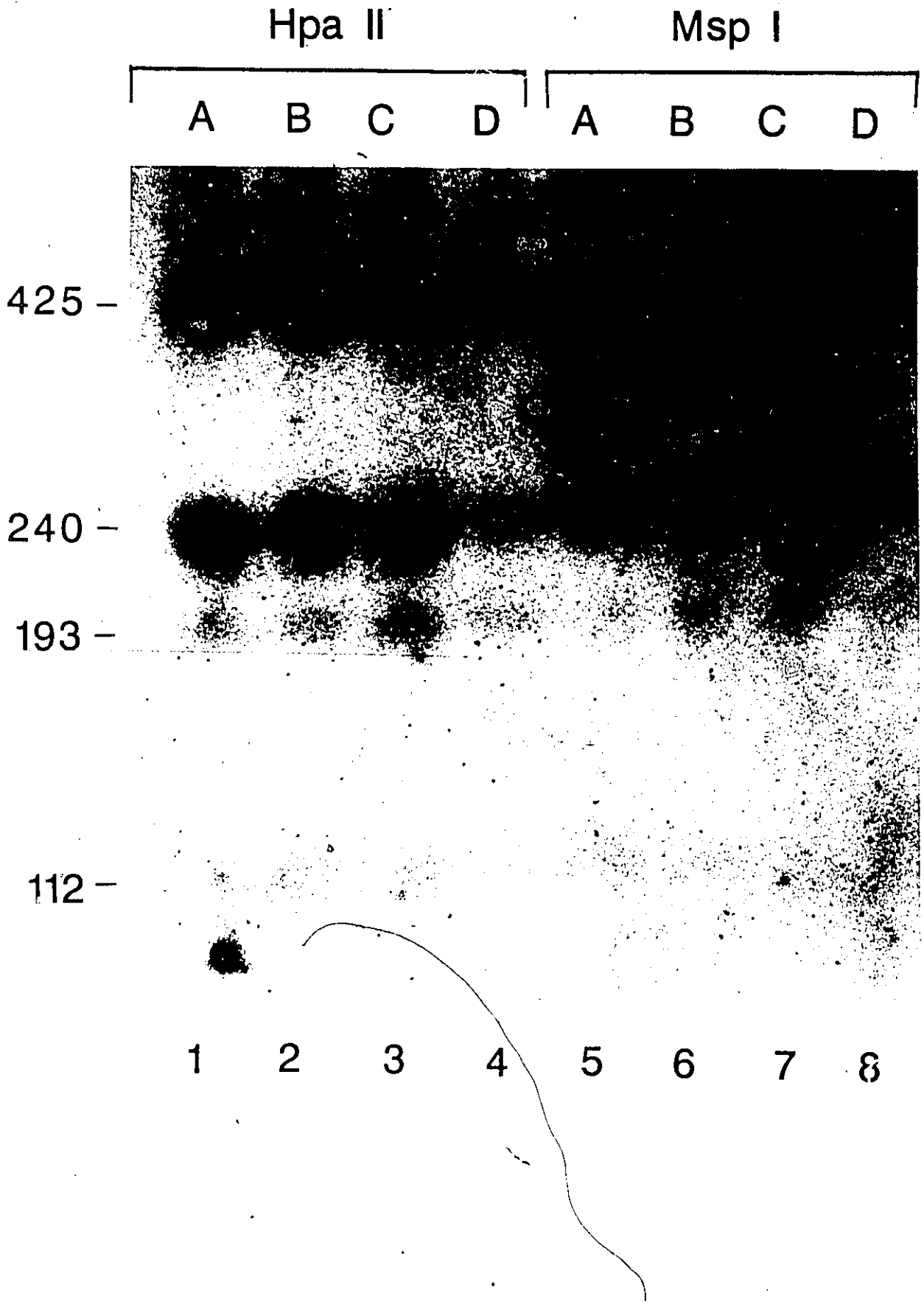

411 –
253 –
215 –
1
2
3
4
5

Fig. 4.4: Lack of methylation of the CCGG sites of the 5' pgk-1 end in undifferentiated, differentiated and reactivated EC cells. Genomic DNA (15 mg) was digested to completion with HpaII (lanes 1-4), or MspI (lanes 5-8) in combination with BamHI and PvuII. Southern blotting and hybridization were carried out as described in the legend of Fig. 4.2. A, AGM2; B, C100; C, DNA from differentiated AGM2 cells; D, DNA from AGM2 reac-2a cells that have been reactivated by treatment with 5-azacytidine. AGM2 reac-2a had higher levels of pgk-1 mRNA than did cells with a single active copy of pgk-1 (A. Hockey, C.N. Adra and M.W. McBurney, manuscript in preparation; also see Appendix 39).


Hpa II
Msp I
A
B
C
D
A
B
C
D
425 –
240 –
193 –
112 –
1
2
3
4
5
6
7
8

4.4 Conclusions

I have investigated the relationship between DNA methylation and gene expression by examining the methylation pattern in the 5'-flanking region of the mouse pgk-1 gene in cells containing only active and active + inactive X chromosomes and on the X chromosome undergoing inactivation using methylation sensitive restriction endonucleases and DNA purified from various EC cells and from mouse tissues. Densitometric scanning analysis indicated that 70-75% of each of the HpaII sites are unmethylated in adult tissues and that these sites are fully unmethylated during the early stages of X chromosome differentiation. In addition, at least four HhaI sites are unmethylated regardless of the state of activity of the pgk-1 gene.

Wolf and Migeon (1982) using two single-copy X chromosome-specific DNA sequences which map to the long arm of the human X, found no differences in DNA methylation patterns between male and female somatic cell DNA. However, the significance of these results is questionable. The assayed sites may not be relevant for gene activity. Also, it is not known if these sites are parts of coding sequences. Moreover, since not all loci on the inactive X chromosome are subject to inactivation, the probes used by Wolf and Migeon may not be part of genes that undergo X inactivation. More meaningful studies of the human (Yen et al., 1984; Wolf et al., 1984) and mouse (Lock et al., 1986) hprt genes, the human g6pd gene (Toniolo et al., 1984; Wolf et al., 1984) and the human pgk-1 gene (Keith et al., 1986)

showed that certain regions of these genes are hypermethylated on the transcriptionally inactive X chromosome. To establish whether the methylation occurs before or after X inactivation, Lock and her co-workers (1987) examined the methylation pattern of 3 HpaI sites of the mouse hprt gene in teratocarcinoma cells and in female embryos at various stages of embryogenesis. They observed that X inactivation preceded the methylation of the hprt gene by a few days. This observation along with the demonstration that DNA demethylating agents are not capable of reactivating the inactive X chromosome in normal human diploid cells (Wolf and Migeon, 1982), and in some embryonal carcinoma cell lines (Paterno et al., 1985) were considered as evidence suggesting that DNA methylation is a secondary mechanism by which the cell maintains or locks the X chromosome in its inactive state. My results are different from those of Lock et al., and indicate that the CCGG sites located in the 5' control region of the mouse pgk-1 gene do not become significantly methylated even a few weeks after the time of X chromosome inactivation. Extensive methylation may not occur at all during the life of the animal.

The studies reported here suggest that mechanisms other than methylation may be involved in maintaining the X chromosome silent. The observation that DNA methylation does not seem to play a role in the repression of the marsupial g6pd gene (Kaslow and Migeon, 1987) is also consistent with this view. In addition, the fact that the oct locus remains silent even after reactivation by treatment with 5-AC (Appendix 1) is strong evidence that mechanisms other than DNA methylation are involved in turning on and off genes.

My results showing lack of methylation in the 5' end of the mouse

pgk-1 gene on the inactive X chromosome are in direct contrast with the results of Keith et al. showing that the human pgk-1 gene inactivity is correlated with the extensive methylation of its 5' end. An important weakness in all of these studies is that only a subset of the CpG sequences present in the 5' end of the mouse pgk-1 gene can be recognized by the available restriction endonucleases.

The lack of conservation of the pattern of differential methylation between the mouse and human pgk-1 genes suggest that the portion of the mouse genome that I have examined may not be relevant to the mechanism of X inactivation. Therefore, it is important to look for methylation differences in and around the pgk-1 gene in our EC cell lines and in the developing female mice embryos. This can be done using field inversion gel electrophoresis (Carle et al., 1986) of DNA digested with enzymes which cut rarely but have a CpG sequence in their recognition site. Mrs. M. Bartlett, a graduate student, is currently addressing this issue.

CHAPTER V

THE FAMILY OF MOUSE PHOSPHOGLYCERATE KINASE GENES AND PSEUDOGENES

5.1 INTRODUCTION

The pgk-1 and pgk-2 genes encode phosphoglycerate kinase (PGK; ATP; 3-phospho-D-glycerate 1-phosphotransferase, EC 2.7.2.3), a key enzyme in glycolysis. In most mammals investigated (VandeBerg, 1985), the PGK in somatic cells is encoded by the X-linked pgk-1 gene while the pgk-2 gene is autosomal and expressed exclusively during spermatogenesis.

To understand the evolutionary relationship between the pgk-1 and the pgk-2 genes I decided to clone the pgk-2 gene. In addition I decided to investigate the mouse genome for the presence of other pgk-related sequences. Our studies revealed that the pgk-2 is a very ancient retroposon derived from pgk-1. In addition to pgk-1 and pgk-2, six other pgk-1-like sequences were present in both Balb/c and C3H/He strain mice. All were located on autosomal chromosomes, and all appear to be retroposed pseudogenes originating relatively recently from pgk-1.

RESULTS

5.2 Mouse Genomic Sequences Homologous to pgk-1 cDNA:

I used three cDNAs as probes in my studies and these are

described schematically in Fig. 5.1. Probe A is a cDNA derived from human pgk-1 mRNA (Michelson et al., 1983). This cDNA clone spans the entire coding region of the transcript. Probe E (a gift of Michelson and Orkin) is a partial cDNA derived from mouse pgk-1 mRNA. This cDNA was sequenced by Dr. P. Boer (Chapter 3) and is identical to nucleotides 918-1639 of the cDNA reported by Mori et al. (1986). Probe G (cloned by Dr. Y.-F. Lau and his co-workers) is a cDNA reactive with pgk-2 mRNA. It contains the entire coding region of the pgk-2 transcript (see section 5.9). Probes B, C, and D were derived from probe A by restriction enzyme digestion and subcloning into plasmids. These probes react with sequences primarily in the 5'-untranslated region (B), the amino terminus of the coding sequence (C), and the middle of the coding sequence (D). Probe F is derived from probe E and is specific for the 3'-untranslated region.

DNA isolated from C3H/He mouse strain was digested with BamHI, electrophoresed in an agarose gel and blotted onto a nylon membrane. When I probed with probe E, seven distinct bands were evident (Fig. 5.2, lane 1). These bands were designated B2, B6, B8, B10, B12.5, B13, and B24 to indicate that they were BamHI fragments of approximately 2, 6, 8, 10, 12.5, 13, and 24 kb in length, respectively.

Fig. 5.1: Phosphoglycerate kinase cDNA probes. A is a 1.9-kb full-length human pgk-1 cDNA derived from the recombinant plasmid pHPGK7-e (Michelson et al., 1983). B, C and D are subcloned fragments derived from probe A using the restriction sites indicated. E is a 0.72-kb mouse pgk-1 cDNA obtained from Michelson and Orkin. F is derived from E. G is a 1.45-kb full-length mouse pgk-2 cDNA derived from the plasmid pMPGK-2 (Boer et al., 1987). The restriction sites indicated are: E, EcoRI; H, HindIII; L, SalI; P, PstI; Ss, SstI; V, PvuII. Open boxes represent untranslated regions; the filled boxes represent coding regions.

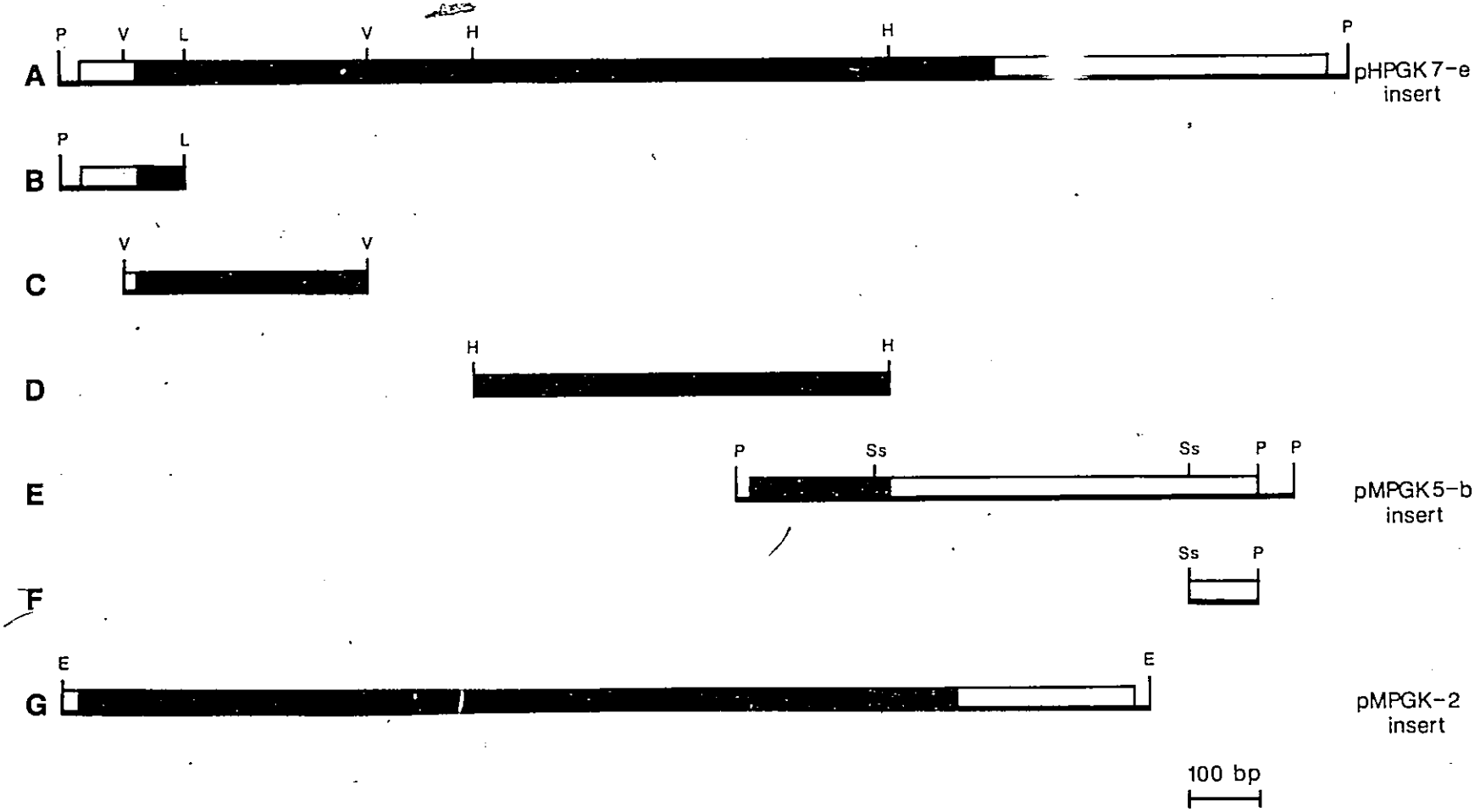
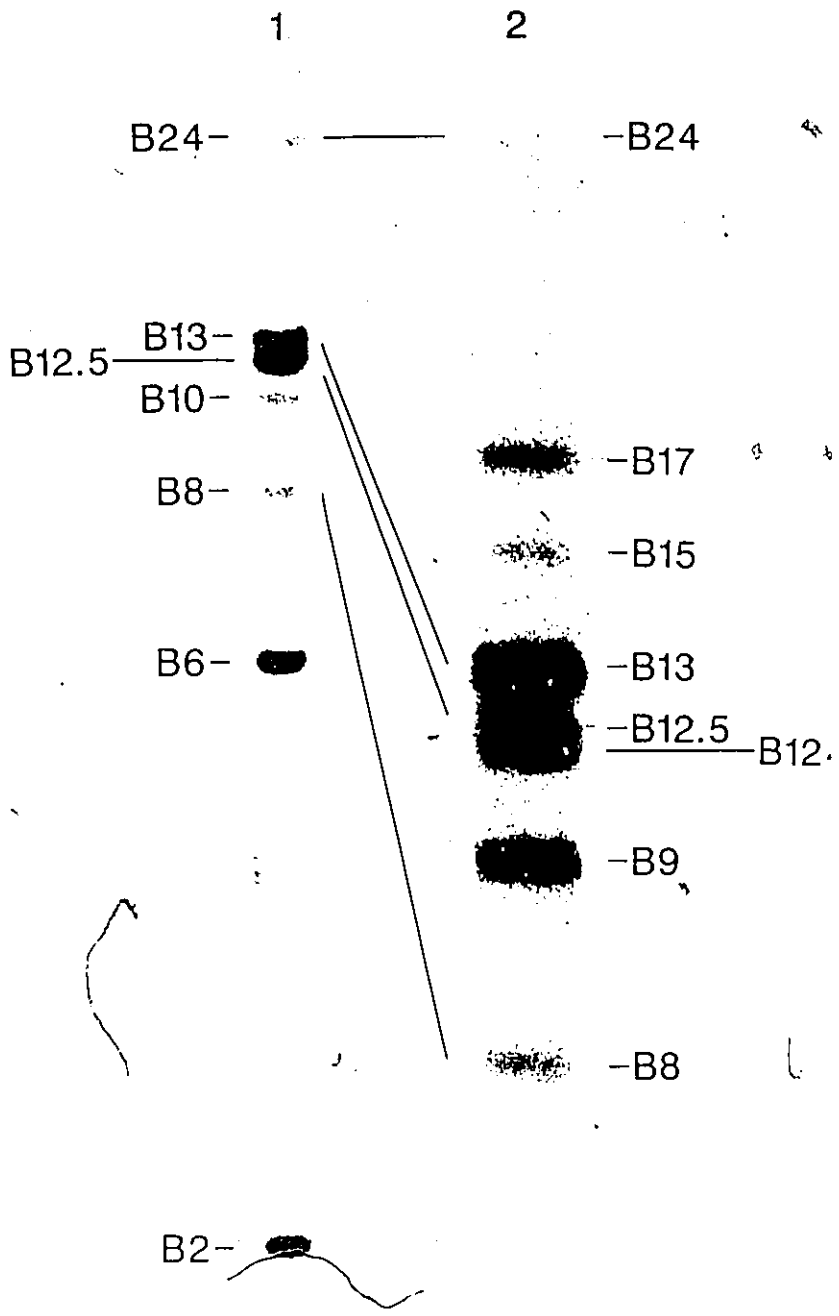

A
P
V
L
V
H
H
P
pHPGK7-e
insert
B
P
L
C
V
V
D
H
H
E
P
Ss
Ss
P
P
pMPGK5-b
insert
F
Ss
P
G
E
E
pMPGK-2
insert
100 bp

Fig. 5.2: Southern analysis of mouse DNA indicates a number of pgk-1 related sequences. High-molecular-weight DNA (10 µg) from mouse spleen was digested to completion with BamHI, electrophoresed in 0.6% agarose, and blotted for hybridization. Lane 1: DNA from C3H/He strain mice hybridized to probe F; lane 2: DNA from Balb/c strain hybridized to probe A. The bands with pgk-related sequence are labelled B, for BamHI, with a number corresponding to the approximate length in kilobases.


1
2
B24–
–B24
B13–
B12.5
B10–
B8–
B6–
–B17
–B15
–B13
–B12.5
B12
–B9
–B8
B2–

Southern blots of BamHI-digested DNA isolated from Balb/c strain mice, when hybridized to probe A, revealed four additional bands of approximately 9, 12, 15, and 17 kb (Fig. 5.2, lane 2). The B12 and B17 bands appeared only in blots hybridized to probe A and were, therefore, presumed to react with DNA sequences present near the 5' end of the cDNA. The B9 and B15 bands appeared only in Balb/c DNA, suggesting that they represent restriction fragment length polymorphisms (RFLPs). The DNAs isolated from C3H/He and Balb/c mice are compared in Fig. 5.3. In lanes 1 and 2, C3H/He and Balb/c DNAs, respectively, were hybridized with probe E. The two strains shared fragments B6, B8, B12.5, B13, and B24. In addition, the C3H/He strain contained fragments B2 and B10, which were absent from DNA of Balb/c, whereas the Balb/c strain contained fragments B9 and B15 absent from C3H/He.

5.3 Mapping RFLP fragments:

To determine the relationship between these strain-specific fragments, I examined them in recombinant inbred strains derived from crosses between Balb/c and C3H/He mice (Taylor, 1981) in collaboration with Dr. N. Ellis. The fragments present in DNAs isolated from a variety of strains are shown on Table 5.1. The B9 and B10 fragments segregated in these recombinant inbred strains, indicating that they are alleles. The pattern of segregation suggests that they are linked to the car gene (encoding carbonic anhydrase) which maps to chromosome 3 (Paul and Elliott, 1987). The B15 and B2 fragments are not alleles because they failed to segregate in the recombinant inbred strains.

Fig. 5.3: Restriction fragment length polymorphisms of the mouse pgk-related sequences. High-molecular-weight DNA (10 mg except in lane 5 where 5 mg was loaded) from mouse spleen was digested to completion with BamHI, electrophoresed in 0.6% agarose, and blotted for hybridization. Probe E was used for hybridization to C3H/He DNA (lane 1) and to Balb/c DNA (lane 2). B15 and B9 were present in Balb/c DNA, whereas B10 and B2 were present in C3H/He DNA. Probe A was used for hybridization to lanes 3-5 containing DNA from a C3H/He female carrying pgk-1^a and pgk-1^b (lane 3), a male carrying pgk-1^b (lane 4), and a female carrying pgk-1^a (lane 5). The B17 fragment correlated with the pgk-1^b allele while the pgk-1^a allele gave rise to the B21 band. Polymorphic bands are indicated with an asterisk.

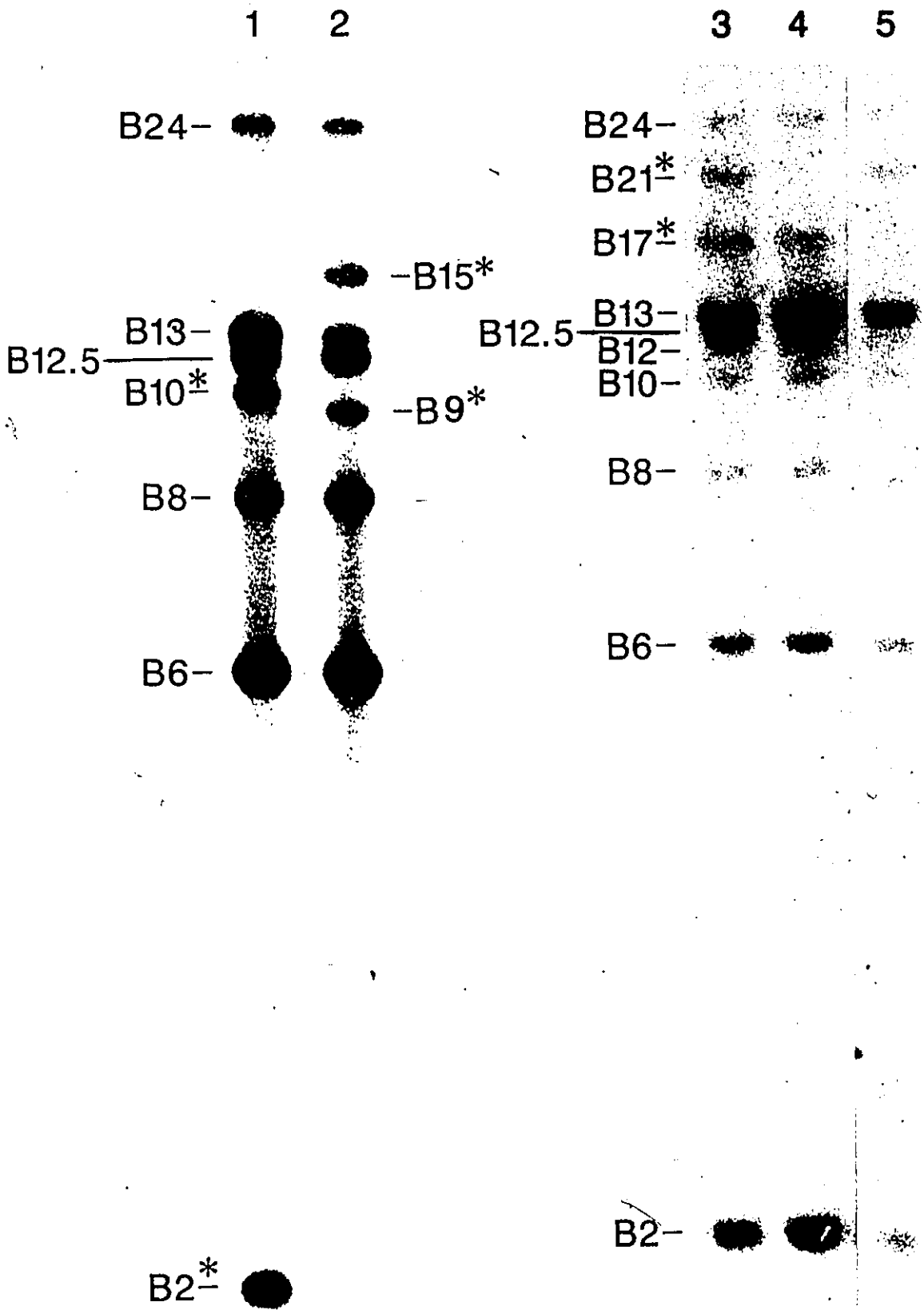

1
2
3
4
5
B24–
B13–
B12.5
B10*
B8–
B6–
B2*
–B15*
–B9*
B24–
B21*
B17*
B13–
B12.5
B12–
B10–
B8–
B6–
B2–

TABLE 5.1: Mapping of pgk-Related Fragments in Recombinant Inbred Mice[a]

	Balb x C3H strain number														
	B	H	2	3	4	6	7	8	9	10	11	12	14	19	Chromosome
B15	+	-	+	+	+	-	-	+	-	+	-	+	-	-	4
B2	-	+	+	-	+	-	+	+	-	-	-	+	+	+	ND
B9	+	-	-	+	-	+	-	+	+	+	+	-	+	+	3
B10	-	+	+	-	+	-	+	-	-	-	-	+	-	-	3

[a]DNA from the various recombinant inbred strains derived from Balb/c and C3H/He mice were obtained from B.A. Taylor (Jackson Laboratory, Bar Harbor, Maine). The segregation pattern of each pgk-related DNA sequence was matched to those of mapped genes in order to establish linkage. B, Balb/c; H, C3H/He. ND, not determined.

The pattern of hybridization of the B15 fragment is most consistent with a location on chromosome 4 proximal to alkaline phosphatase 2 (Wilcox and Taylor, 1981) and distal to the ornithine decarboxylase pseudogene marker BRS-4 (Richards-Smith and Elliott, 1984). The B2 fragment appears to be linked to plasma cell antigen 1 (B.A. Taylor, unpublished results), which has not yet been mapped.

I detected one additional polymorphism with probe A and it is shown in Fig. 5.3, lanes 3, 4 and 5 (also see Chapter III). The DNA in these three lanes came from congenic strains of C3H/He animals containing different alleles encoding PGK-1. As previously noted by S. Berger and V.M. Chapman (personal communication) the pgk-1^b allele appears to segregate with the B17 fragment, while a larger fragment, B21, segregates with the pgk-1^a allele. Breeding studies by Berger and Chapman mapped the B17 and B21 fragments to the X chromosome. I compared the intensities of the pgk-related fragments in DNA isolated from male and female mice (Adra et al., 1987) and showed that B17 and B12.5 were twice as intense in female as in male DNA, indicating the location of these two fragments on the X chromosome. All other fragments were equally intense, indicating their presence on autosomal chromosomes.

5.4 Hybridization at elevated stringency:

The hybridization shown in Figs. 5.2 and 5.3 was performed under conditions normally described as high stringency. To more closely examine the degree of homology between the cDNA probe E and the genomic sequences, hybridization was performed at elevated temperatures (Gusella et al., 1982). The results (Fig. 5.4) indicate

that specific hybridization to fragments B6, B12.5, B13, and B24 was retained even under very high stringency conditions, whereas hybridization to the B8 and B10 fragments was eliminated. The sequences present in B8 and B10 are, therefore, less homologous to the probe than those present in the other four fragments.

5.5 Cloning of Mouse pgk sequences:

Probes A and E were used to isolate all pgk-related sequences from genomic libraries of Balb/c strain mice. The B24 fragment was isolated by Dr. N. Ellis from a library of partially MboI-digested DNA cloned into the cosmid vector, pHC79 (Hohn and Collins, 1980). The B6, B8, B10, B12, B12.5, B15, and B17 fragments were isolated from libraries that I made from DNA digested to completion with BamHI and cloned in the bacteriophage lambda vector, EMBL3 (Karn et al., 1980; Frischauf et al., 1983). The inserts from recombinant lambda genomes were characterized based upon fragment size, and a representative of each was subcloned into the plasmids pSP64 and pSP65 (Melton et al., 1984) for restriction mapping (see Appendix 2). The restriction maps are shown in Fig. 5.5.

Fig. 5.4: Hybridization of Southern blots at elevated temperatures. Each lane contained 10 mg of DNA from spleen of male C3H/He mice digested with BamHI and electrophoresed in 1% agarose gels. The blotted filters were hybridized for 16 h with the same preparation of radiolabelled probe E at 42°C (lane 1), 50°C (lanes 2,3,6,7), and 55°C (lanes 4,5,8,9). All filters were washed at 60°C in 0.2 x SSC and 0.1% SDS. The X-ray film was exposed for 32 h (lanes 1-5) or 10 days (lanes 6-9).

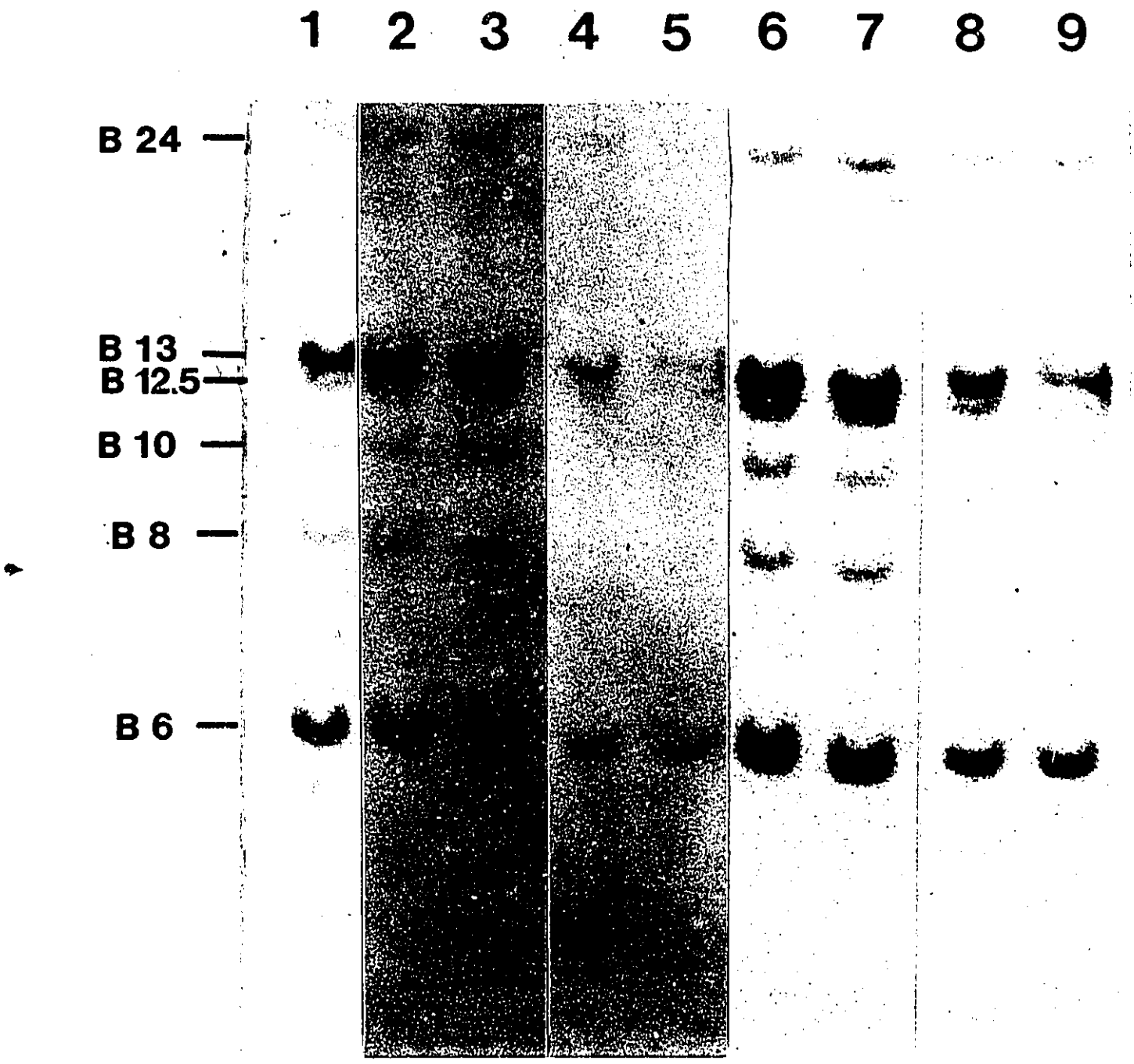
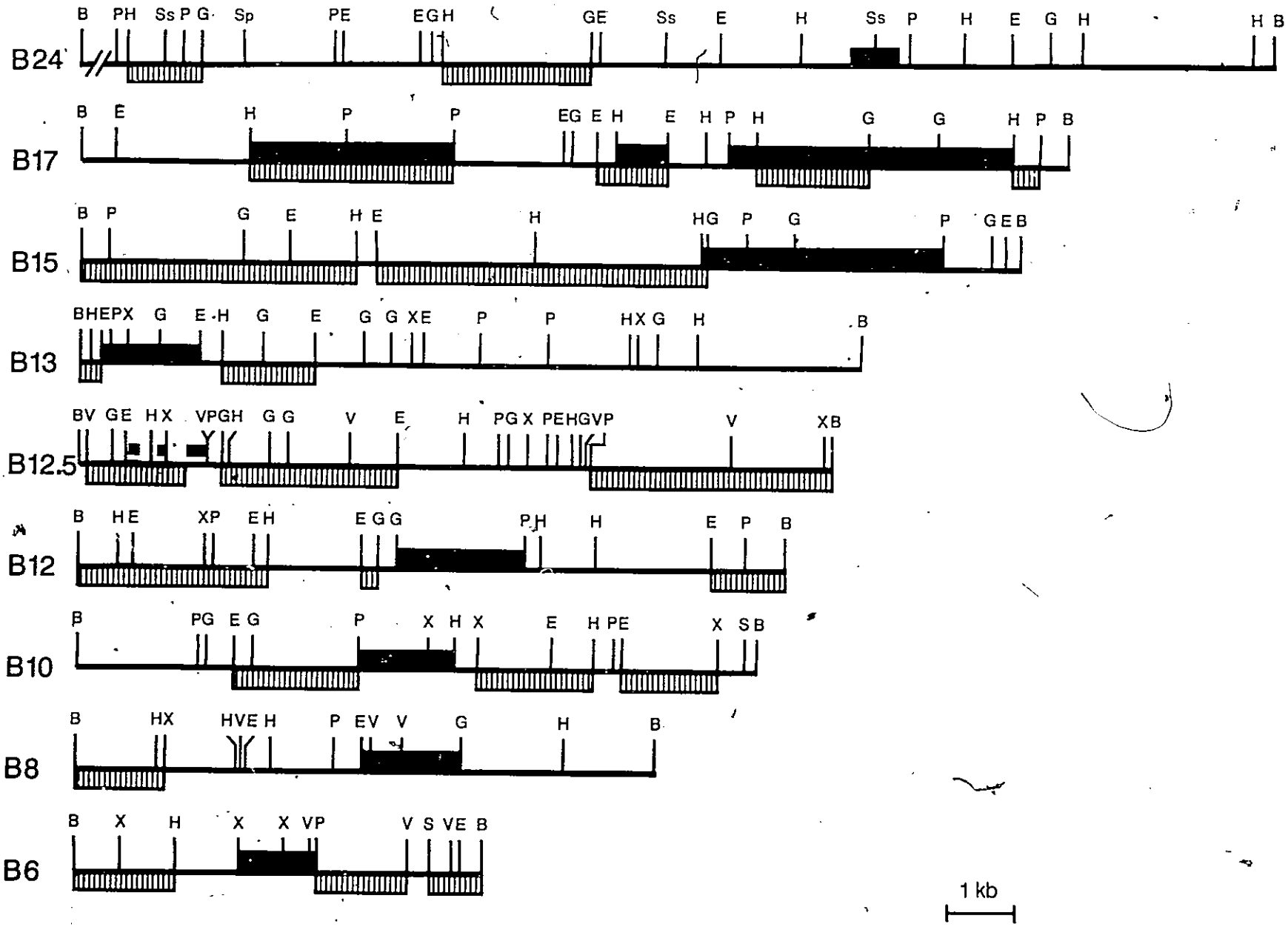

1 2 3 4 5 6 7 8 9
B 24
B 13
B 12.5
B 10
B 8
B 6

Fig. 5.5: Restriction maps of the mouse genomic fragments homologous to pgk-1 cDNA probes. The recombinant bacteriophage or cosmid clone inserts were subcloned into pSP64 or pSP65 plasmids for restriction mapping. Those restriction fragments that hybridized to probe A are indicated as solid boxes above each line; those that hybridized to total genomic mouse DNA are shown as hatched boxes below each restriction map. The restriction enzymes used were: B, BamHI; E, EcoRI, G, BglII, H, HindIII, P, PstI, S, SalI; Sp, SphI; Ss, SstI; V, PvuII; X, XbaI.


B24
B PH Ss P G Sp PE EG H GE Ss E H Ss P H E G H H B
B17
B E H P P EG E H E H P H G G H P B
B15
B P G E H E H HG P G P G E B
B13
B H E P X G E H G E G G X E P P H X G H B
B12.5
B V G E H X V P G H G G V E H P G X P E H G V P V X B
B12
B H E X P E H E G G P H H E P B
B10
B P G E G P X H X E H P E X S B
B8
B H X H V E H P E V V G H B
B6
B X H X X V P V S V E B
1 kb

Each cloned fragment had a different restriction map indicating that each of the nine fragments is unique. Those restriction fragments that hybridized to probe A are indicated as black boxes above the line and those restriction fragments that hybridized to nick-translated total mouse genomic DNA appear as striped boxes below the line. These latter fragments are presumed to contain repetitive sequences.

In general, restriction fragments that hybridized to probe A did not hybridize to total nick-translated DNA, indicating that the repeated sequences are normally not intermingled with unique sequences homologous to the pgk-1 cDNA. The exceptions occur in fragments B17 and B12.5, in which many of the restriction fragments hybridized efficiently to both total genomic DNA and pgk sequences. These two fragments comprise the pgk-1 gene (Chapter 3). Thus it appears that many of the introns in the pgk-1 gene contain repeated sequences, but none of the other pgk-related sequences are similarly interrupted by such intron-like regions.

5.6 Hybridization to Regions of pgk-1 cDNA:

Since only two genes encoding PGK proteins are known in mouse, many of the nine fragments cloned are likely to be pseudogenes derived from the pgk-1 or pgk-2 genes. In order to further characterize the genes and pseudogenes, I made use of the seven probes outlined in Fig. 5.1 and hybridized each to the cloned fragments reported in Fig. 5.5. The results are summarized in Table 5.2.

The B12 fragment hybridized strongly to the pgk-2 cDNA (probe G)

but not to probes for the 5' (B) or 3' (E, F) ends of the pgk-1 cDNA. This fragment contains the pgk-2 gene (see section 5.9).

The B17 fragment hybridized to all probes except that specific for the 3'-untranslated region, while B12.5 hybridized only to probes containing sequences from the 3' end of the pgk-1 cDNA. These fragments do, in fact, contain the 5' and 3' ends of the pgk-1 gene, respectively (Chapter 3).

The fragments B13 and B15 hybridized to probes A-F, indicating that they contain sequences homologous to all regions of the pgk-1 cDNA, including coding and noncoding regions. B24, B10, B8, and B6 hybridized to only a subset of those probes spanning the pgk-1 cDNA. Of these latter four fragments, only B6 hybridized to probe F and only B10 hybridized to probe B. Thus, if these four fragments do represent pseudogenes, each has been truncated at either its 3' end (B10), its 5' end (B6), or both ends (B24, B8).

TABLE 5.2: Hybridization of Genomic Fragments to cDNA Probes

Genomic Fragment	Probe A	B	C	D	E	F	G
B24	+	-	-	+	+	-	-
B17	+	+	+	+	weak	-	+
B15	+	+	+	+	+	+	-
B13	+	+	+	+	+	+	-
B12.5	+	-	-	weak	+	+	-
B12	+	-	+	weak	-	-	+
B10	+	weak	+	weak	+	-	-
B8	+	-	-	weak	+	-	weak
B6	+	-	-	+	+	+	-
B2	+	ND[a]	ND	ND	+	ND	-

[a] ND, not determined.

5.7 Sequence of B24 Pseudogene:

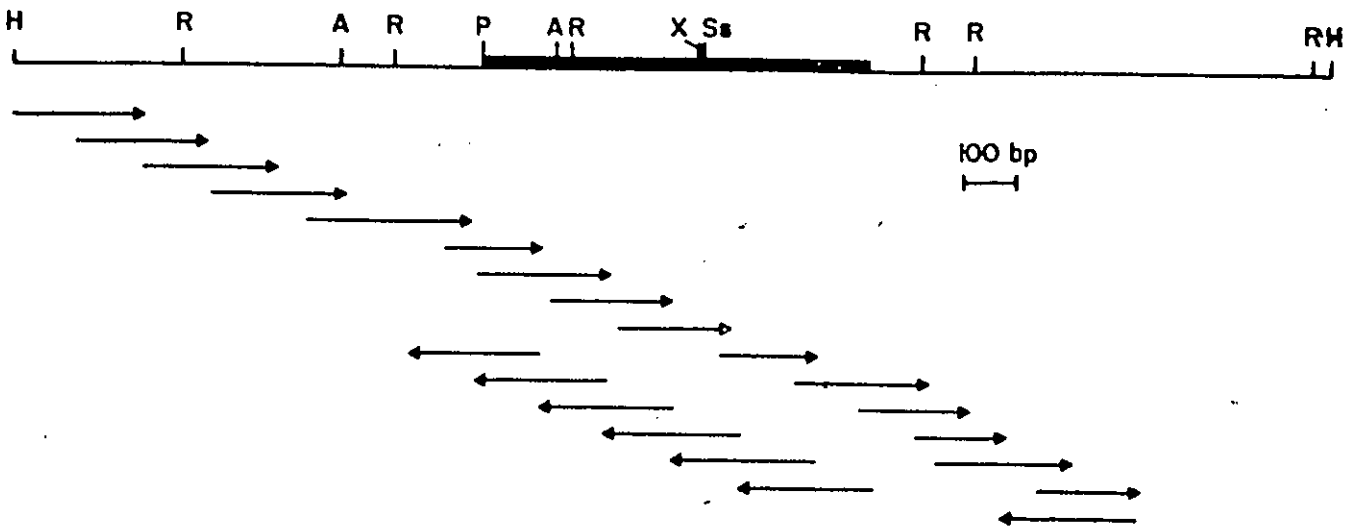
The B24 fragment contained a sequence highly homologous to the pgk-1 cDNA coding region (Fig. 5.4) but not homologous to either non-coding region. The hybridizing region is contained within a 2.5-kb HindIII fragment that was subcloned into the bacteriophage M13-mp19 for sequencing by Dr. N. Ellis using the strategy outlined in Fig. 5.6.

The sequence is also presented in Fig. 5.6. A region of 710 nt aligns with a high degree of homology to the sequence of mouse pgk-1 cDNA (Mori et al., 1986). A number of stop codons are present in the B24 sequence, indicating that no protein product could be translated from this sequence.

The region of homology begins in the coding region corresponding to codon 263 and extends 230 nt into the 3'-untranslated region but more than 150 nt upstream of the normal pgk-1 polyadenylation site (Mori et al., 1986). The mouse and human pgk-1 genes have 11 exons (Chapter 3; Michelson et al., 1985). The intron-exon borders for the mouse gene have not all been determined but, assuming that they correspond to those of the human pgk-1 gene, the region of homology between the B24 sequence and pgk-1 starts 35 nt into exon 8. Thus the borders of homology between the B24 pgk region and the mouse pgk-1 cDNA do not correspond to intron-exon borders, and introns 9 and 10 present in the pgk-1 gene are absent from the B24 sequence.

The region of B24 homologous to the pgk-1 cDNA ends abruptly at nt 900 and 1609; however, the sequences flanking the ends of the pseudogene lack flanking direct repeats and an A-rich 3' end. A number of unusual sequences are indicated in Fig. 5.6. The 11-nt

Fig. 5.6: DNA sequence of the B24 pseudogene. The 2.5-kb HindIII fragment of B24 was subcloned into M13-mp19 and serial deletions constructed to give a sequencing strategy as outlined in upper panel. Restriction sites are indicated for: A, AccI; H, HindIII; P, PstI; R, RsaI, Ss, SstI; X, XbaI. The sequence is shown in lower panel. The region between bases 900 and 1610 was highly homologous to the mouse pgk-1 cDNA sequence of Mori et al. (1986). Base identity in this region is indicated by a star, base differences are shown in lower case, and gaps are indicated. The underlined region from 1965 to 2014 consists of 50 bases of alternating (AN) where N is C or G. The underlined region from 2026 to 2070 contains nine copies of the pentamer, AGAGG. There were no direct repeats on either side of the pgk-related sequence, but the boxed 11-mer at positions 577 and 1605 have identical sequences.




```
1    AAGCTTGGTCAGAAGTATTTTACACAGTTGGATCATTCTGCCACTTTTCTGTTTAATGTGGCCATTTCCTGGCCCTGTTGTAATTCAGTAAGGTTTAGTGAGAGCAGCCCCTCATTTTAT
121  CCAATTTTGCCAGTTTGTGGTTTTCTCTATTATATTATGTTGTTGCTGCCAATTATAATTTGCTCTGTCTTTTTCGTTTTTTTCCTGTGTATTGTTAACTAGAACCCCCACCTCTCAAAT
241  CATCCATTGCACTGAGCATATAATTGCTTCTCTTACTGAATGCAGGAGAGAAATCTAAGTAGCACAGAAGGAATGCTCTTGTACTCTCCTTTCTAGCAATTTTCTGTTGTTTTTCATGTC
361  GCATCGTGTGATTTGTAATATTGTGTCTAATATAGTTTTCATTATCCTTTCTTCTCCCTTTATTTGTTTGATTTAGTTGTTTTTATATGTAGCTATAATGGAAGTTTAAACATTCTGAAT
481  GGCATAATGGCACATACTATTTTAATTTCAGCTTCGTGTCTCGTATTCCAAAATGCAGGTCAAAAGAGTCTTAGTAAACACAATCAAATTTCCACCATATTTTGACAGTTTTTATAAGAT
601  CAAGTTCACCTTTGAAATGTATACACCATATGCATATACCATATGTGTAAACTATAGATTGTATAGTATAGTTGTGATTTGGTTTACGGTATATTTTCTTTGTGTGGGTATTTTGTTAGG
721  GTACAGTATGTGCATCCATGCAAATTAGGTAAATGTCAAAGTAAGATTATAGGCTTCATTACTCATTTTCTTCTTTTTTACCATGAGAAAGTATCTCTCAGGATCTTCCCTCCGCCCCCC
841  CACCCCGTGGACATCTTTATTGTTTAATAGATCATTAACTTCTGCAGACTTATAGCTGGGATCGTCAAGGATCTCATGTCCAAAGCCGAGAAAAATGGTGTGAAGATTACCTTGCCTGCT
                                agg***caa***t*****a**********************t*****************************t*
961  GACTTTGTCACTGCTGACAAATTTGATGAGAATGCCAAGACTAGCCAAGCTACTGTGGCCTCTGGTATAGCTGTTGCCTGCATGGGCTTGGACTGTCGTACTGAGAGCAGCAAGAAATAT
     ****************************************g**********************************c*****************************************
1081 GCCGAGCCCGTGGCTGAGCTAAGCAGATTGTTTGGAATGGTCCTTTTGGGTTATTTGAATGGGAAGCCTTTGAAAGGCGAACCAAGTCACTCATGGATGAGGTCGTGAAAGCCTCTTCTA
     ********t****g**********************************g*****g*********************cc*************************************a******
1201 GGGGTTGCATCATTATCATAGGTGGTGGAGACACTGCCACTTGCTGTGCCAAATGGAACACAGAGGATAAAGTCAGCCATGTGAGCACTGGGGGCAGTGCAGTCTAGAGCTCCTGGAAGG
     ************c*************************************************************************************g************************
1321 TAAAGTCCTTCCTGGGGTGGATGCTCTCAGCAATATTGAGTTTTTTTCTTTCCTGACTTTGGTTCCTGTGCACAGCCCCTAAGTCAACCTAGTATTTTCCGCATCTCCATTTGGTGTTAG
     **********************************g**t***a**** ********c****************t*  *t******ct*****gc*g******a*********************
1441 CACAAGATTCAGCTAGTGGCTGAGATGCGGCACCAGGAACCCTTAAATAGTTGCACAGCATCTCAGCTCATCTTTACTGCATCAGGATTTGTCTGCGTTCTTCAAGATCCCATTTAAATT
     *g*************************t***** **  * ****** c*** *** ************g**** ********** *  ** *** ***********************
1561 CCTTAGTGACTAAAACCATTGTGCATTGTAGAGAGCGTCTATTTATATTTTGACACTCTCAGTGATCTTTGTCGATTGCTAGTTTATGACTTCCTCGATTGCCCAGCGAACACTTACGAT
     **********************************g****************c**c*tgag
1681 CTAATATCTCAAACACCAAGAAAGGAAGGGCTC*CTTGTACACCCAGCCATACTAGACAATAGCAAAGGTGCTGGAGATTTGGCTAATGTCTTCATGTTTGTAGAGGATAAGTTATATCA
1801 GCTGAAACATCTGCCAAGTACTTATAATTTGTCTTCTTTAATAAACTGAGTGTAGGTTTATATTGGAGGGAGAGGATGAGGCATGGAGGGAAGAAGAAAATTAAGGAGATAAAATAAACA
1921 TACACAGATACATGGGGGGAGAGACACAGAAGAGACAGACAGAAAGACAGAGACAGAGACAGAGACACAGACACACACACACAGAGAGAGAGGGGGGGAGGAGAGAGGAGAGGAGAGG
2041 AGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGATTTGTAATCTAAGAGCTCGGAAGAGAAGGAAAGACTTGTTCTAGTGCTTGCAGAGTCCAAGGT
```

sequence spanning the 3' end of the pseudogene is repeated at nt 582. This 11-mer is very reminiscent of a direct repeat but for the fact that the upstream 11-mer is not adjacent to the pgk-1-related sequence. Downstream of the pseudogene are two more unusual sequence regions. A 50-nt region from positions 1965 to 2014 consists of $(AG/C)_{24}$, and from nt 2026 to 2070 there are nine perfect repeats of the pentamer, AGAGG.

5.8 Mapping of pgk-like Sequences in Somatic Cell Hybrids:

The two fragments B24 and B12 were not polymorphic in the two strains of mice investigated, but their chromosomal location was determined in collaboration with Dr. N. Ellis using a panel of mouse-hamster hybrids with reduced numbers of mouse chromosomes. Unique fragments derived from B24 (Appendix 2) and from B12 (Appendix 2) were used to probe the DNA from the mouse-hamster hybrids for the presence of the expected BamHI fragment. The results are reported in Table 5.4 and indicate that B24 maps to chromosome 16 and B12 to chromosome 17.

TABLE 5.3: Chromosomal Assignment of Fragments B24 and B12 Using Somatic Cell Hybrids[a]

	Clone of mouse x Chinese hamster hybrid												
Chromosome	3-16	2-3A	3-13	3-12	1-7B-4	6-25	1-3A-3	1-3A-2	3-23	6-6	ADCT-25	R44-1	R44
1	(+)		+		+		+						
2	+	+			+		+	+	(+)				
3	+	+				+			+			(+)	
4		+					+	+					
5	+			+					(+)				
6	+	+		+									
7	+		+			(+)		+					
8	+	+						+	+				
9	+		+										
10		+			+	(+)							
11													
12	+			+				+	+				
13	+			+			+	+					
14	+	+		+	+				+				
15	(+)	+		(+)	+		+	+	(+)				
16	+	+	+					+					
17	+	+		(+)								+	+
18	+	+	(+)		+		+	+	+			(+)	+
19	+				+		+	+					
X	+	+	+	+		+			(+)	+			
T[2;16]						+					+		
T[16;2]										+			
Fragment													
B24	+	+	+	-	-	+	-	+	-	-	+	ND	ND
B12	+	+	-	+	-	-	-	-	-	ND	ND	+	

[a]DNAs isolated from a number of mouse-Chinese hamster somatic cell hybrids were the gifts of Drs. D.R. Cox (Joyner et al., 1985) and D. Pravtcheva and P. D'Eustachio (Smiley et al., 1978). The mouse chromosomes present in each hybrid are indicated, where (+) indicates that only part of the chromosome was present. The presence of the B24 and B12 fragments was assessed in each BamHI-digested DNA preparation using as hybridization probes unique fragments subcloned from B24 (a 1.7-kb HindIII fragment to the right of the pgk-related sequence) and from B12 (a 1.4-kb HindIII/EcoRI fragment to the left of the pgk-related sequence) (see Fig. 5.5).

5.9 Cloning the mouse pgk-2 cDNA

To isolate the mouse pgk-2 cDNA and the pgK-2 gene I initiated in 1985 a collaboration with Dr. Y.-F. Lau. He constructed a cDNA library in the ggtl1 expression vector by purifying RNA from the testes of BALB/c mice. The library was screened by K. Chan and M. Elish with a rabbit antibody specific for PGK-2 protein (Lee et al., 1980). A clone named pMPGK-2 was identified. For further analysis, I transferred the cDNA insert into the sequencing plasmid pSP64 (Melton et al., 1984) and into the pSV2 expression vector (Mulligan et al., 1980). These two constructs were named pCAMPGK-2-64 (Fig. 5.7) and pCASV2pgk-2 respectively. I determined that the size of the insert was approximately 1.6 kbp (see Fig. 5.13B for its restriction map). When I used probe G (Fig. 5.1) to probe mouse genomic DNA only the B17 and B12 fragments hybridized (Table 5.2). When the 180-bp RsaI-BglII fragment derived from the noncoding 3' end of pgk-2 mRNA was used to probe Southern blots of mouse genomic DNA (Fig. 5.8A) in collaboration with P. Boer, a single restriction fragment hybridized in each sample of digested DNA. My studies have shown that B17 is X-linked and that it contains the 5' end of the pgk-1 gene. Therefore the detection of B12 with probe G and the detection of a single restriction fragment of a size expected from the restriction map of B12 indicate that the pgk-2 gene is present as a single copy in the mouse genome.

Using the RsaI-BglII fragment (Fig. 5.8B) as a probe, P. Boer examined the transcription of the pgk-2 gene in different mouse organs. A hybridization signal was detected in testes. This probe did not hybridize to RNA purified from several somatic tissues.

Fig. 5.7: The construct pCAMPGK-2-64 contained the pgk-2 cDNA insert and it was used in restriction analysis and sequencing. The pgk-2 cDNA insert was derived from ggtl1 and ligated into an EcoRI-digested pSP64. Open boxes represent untranslated regions; the filled box represents coding region. E,EcoRI.

pCAMPGK-2-64

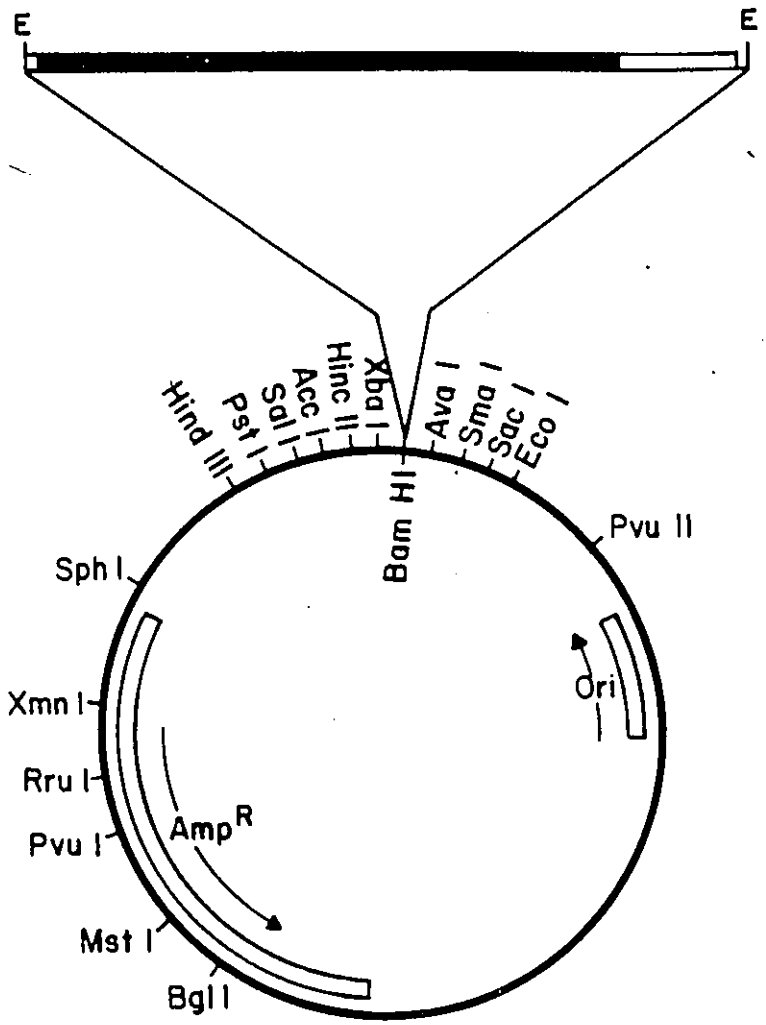

E
E
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Bam HI
Ava I
Sma I
Sac I
Eco I
Pvu II
Sph I
Xmn I
Rru I
Pvu I
Mst I
Bgl I
Amp^R
Ori

Fig. 5.8: The pgk-2 gene is unique in the mouse genome and is expressed only in testes. (A) Southern blot analysis of mouse DNA hybridized with the subcloned RsaI-BglII fragment from the 3' noncoding region of the pgk-2 cDNA (Fig. 5.3A). The DNA was digested with BamHI (lane 1), EcoRI (Lane 2), and HindIII (lane 3); the indicated sizes are in kilobase pairs. (B) RNA blot analysis of different adult mouse tissues: brain (lane 1), testis (Lane 2), and spleen (lane 3); hybridization was with the same probe used in panel A. RNA isolated from heart and skeletal muscle also failed to hybridize the pgk-2-specific probe. (C) The RNA blot in panel B was stripped and hybridized with a subclone of pgk-1 cDNA corresponding to the 3' noncoding region. The positions of mouse ribosomal RNAs (4.71 and 1.87 kb, respectively) are shown on the right. The arrows indicate the positions of the PGK-1 mRNA (top arrow) and the PGK-2 mRNA (bottom arrow).

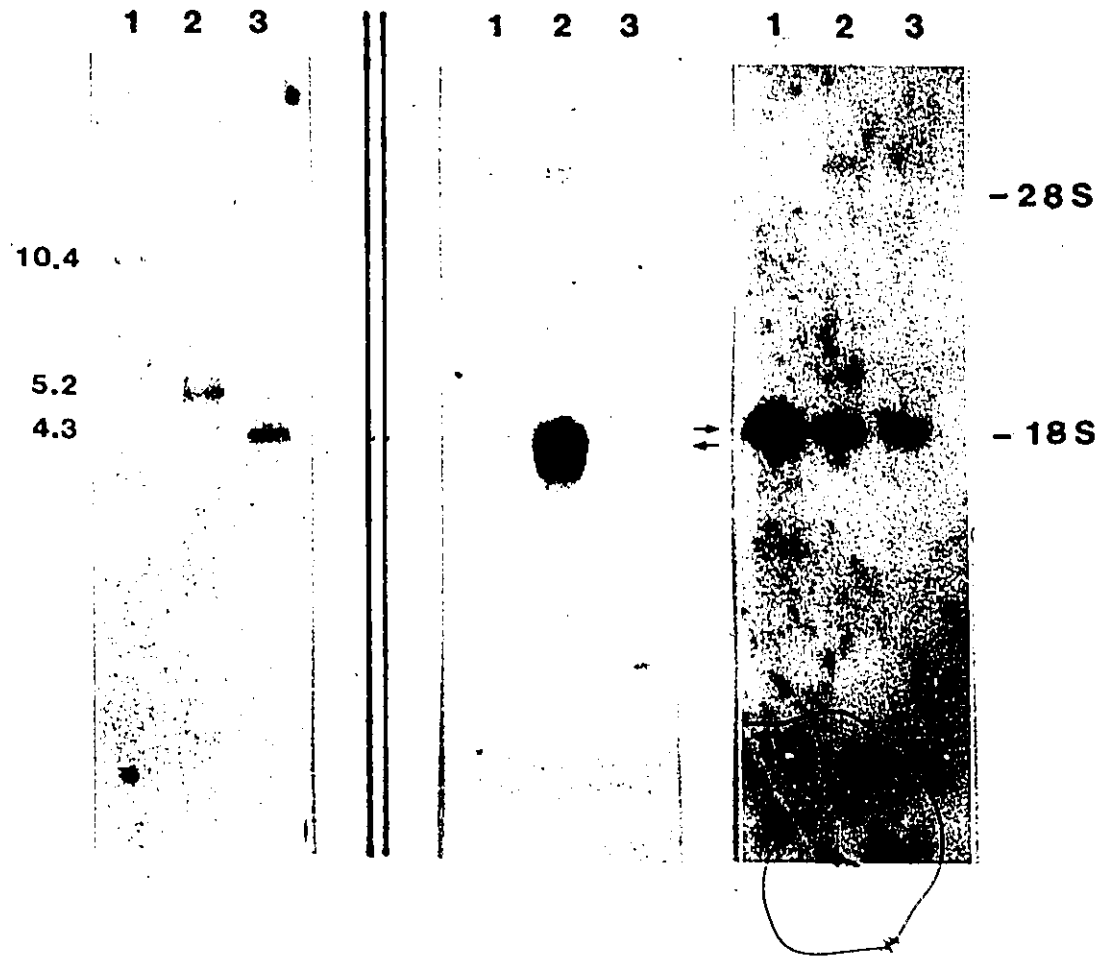
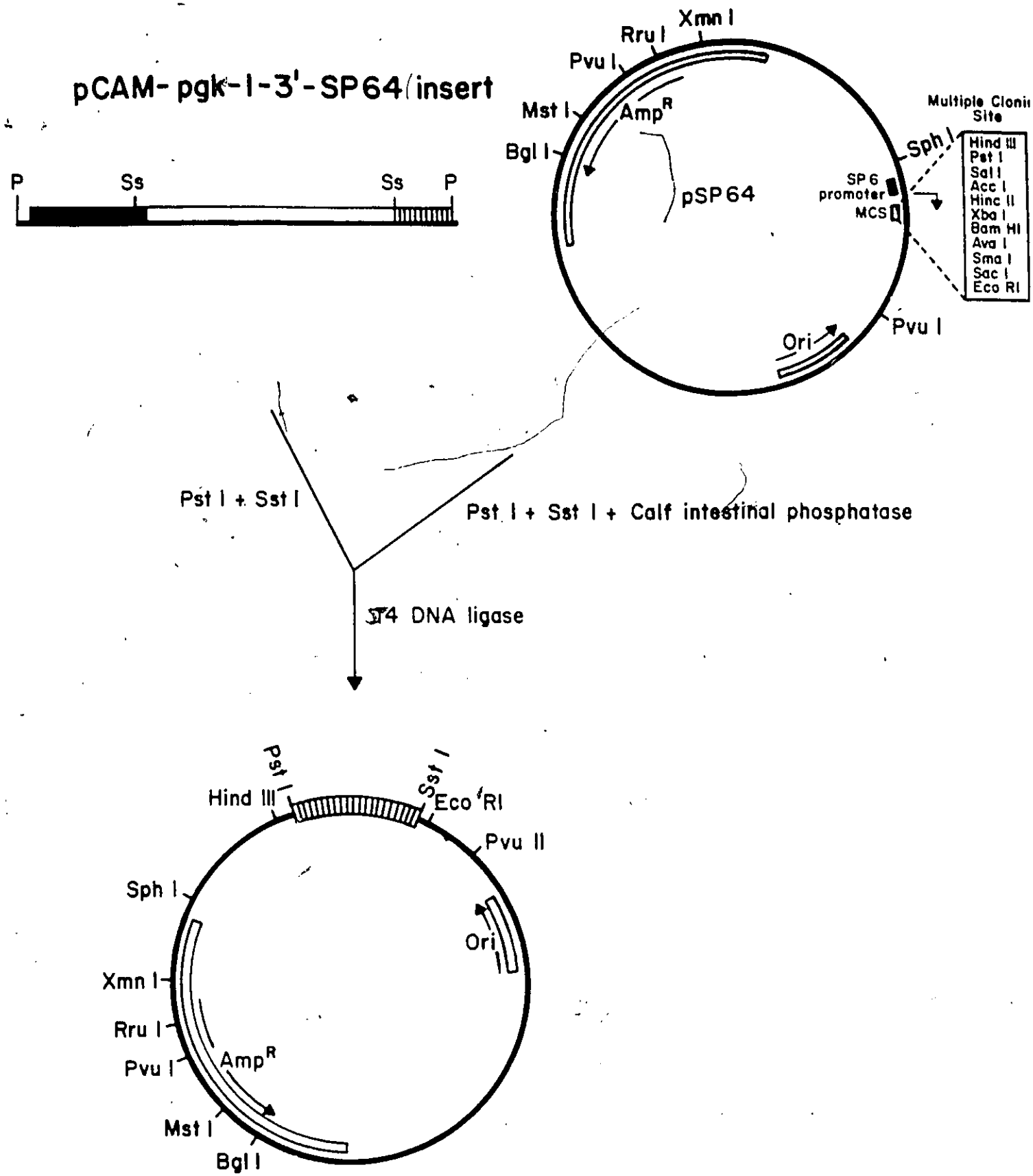

1 2 3
1 2 3
1 2 3
10.4
5.2
4.3
-28S
-18S

Fig. 5.9: The recombinant pCA-M-pgk-1-3'PstI/SstI 100-SP64 used for RNA blot analysis. It contains an approximately 100-bp PstI/SstI fragment homologous to the untranslated 3' end of the pgk-1 mRNA. This construct was prepared by ligating the PstI/SstI fragment (striped box) derived from the mouse pgk-1 cDNA (see Appendix 2) into a PstI/SstI-digested pSP64. Open box represents untranslated region; the filled box represents coding region. P,PstI; Ss,SstI.



pCA-M-pgk-1-3' Pst I/Sst I 100-SP64

A recombinant (Fig. 5.9) containing 100-bp SstI/PstI fragment homologous to the noncoding 3' end of the pgk-1 mRNA was also used to probe the same Northern blots. A hybridization signal at about 1.8 kb was detected in all organs (Fig. 5.8C).

In collaboration with Dr. M.W. McBurney the expression vector pCASV2pgk-2 was transfected into Hela cells. This vector directed the synthesis of the mouse Balb/c isoenzyme of pgk-2 (Fig. 5.10, lane 1). These results suggest that the 1.6 kbp pMPGK-2 insert contains all the coding sequence for PGK-2 protein.

5.10 Cloning the pgk-2 gene:

Based on the above results: a) the strong hybridization of probe G to B12 and the pattern of hybridization of B12 to the various regions of the pgk-1 cDNA (Table 5.2) and b) the mapping of B12 to chromosome 17, I concluded that B12 is the only fragment amongst the cloned mouse pgk sequences (see Fig. 5.5) likely to contain the pgk-2 gene. To confirm that the B12 fragment (Fig. 5.3A) comprised the mouse pgk-2 gene, I cloned it in pSP64 (Fig. 5.11). This construct was named pCAB12-SP64. In collaboration with Dr. M.W. McBurney, this construct was cotransfected with pSVneo (Southern and Berg, 1982) into NIH 3T3 cells. NIH 3T3 cells were also transfected with pSVneo alone as a control. The cells were selected in G418. 25 resistant colonies were pooled, extracts prepared, electrophoresed in cellogel strips, and stained for PGK enzyme activity (Bucher et al., 1980; Paterno and McBurney, 1985). The NIH 3T3 cells transfected with the pCAB12 plasmid contained a PGK enzyme activity with mobility similar to that

of enzyme found in testes of C3H/He mice (Fig. 5.10, lanes 3 and 6). The cells transformed with pSVneo did not contain any PGK-2 activity (Fig. 5.10, lanes 4 and 7). The expression results established unequivocally that the plasmid pCAB12 does carry the mouse pgk-2 gene.

By using probe A, I have shown that the B12 fragment contained a sequence homologous to the pgk-1 and the pgk-2 cDNAs (Fig. 5.5). By restriction mapping and hybridization studies I determined that the hybridizing region is contained within a 2.1-kbp PstI-BglII fragment. I subcloned this fragment into the plasmids pSP64 and pSP65 for further restriction mapping and sequencing. These two constructs were called pCAB12Bgl/P2.1-SP64 and pCAB12Bgl/P2.1-SP65 respectively (Fig. 5.12). I subjected both the pgK-2cDNA and the genomic fragment to restriction digestion and electrophoresis. The restriction map of the pMPGK-2 insert was identical to that of a large segment of the PstI-BglII genomic fragment (Fig. 5.13). These results indicated that the pgk-2 gene appears to be an intronless gene. (In fact these studies were done before those described in Fig. 5.8.)

I suspected that B12 carries the pgK-2 gene because of the hybridization studies (table 5.2). B12 was the only fragment that did not react with probe E.I felt that it is unlikely that B12 is a 3'-truncated pgK-1 pseudogene. I attributed the lack of hybridization to sequence divergence at the 3'-end between the pgK-1 and the pgK-2 genes. However several scientists refused to believe that B12 is anything else but a piece of "garbage" DNA. Their argument was: the region of B12 homologous to the pgK-1cDNA does not seem to be interrupted by introns therefore it is not a gene! In 1977, the scientific world discovered that genes are interrupted. How quickly the concept of split genes became a dogma in some minds! The restriction map led to a change in their attitude.

Fig. 5.10: PGK-2 activity in cells transfected with the pgk-2 gene and cDNA. The full-length pgk-2 cDNA insert of 1.6 kbp was inserted into the pSV2 vector, and the resulting plasmid, pCASV2pgk-2, was transfected into HeLa cells. After 48 h, a cell extract was prepared and run in lane 1. NIH 3T3 cells were transfected with pSV2neo alone (lanes 4 and 7), pCASV2pgk-2 (lanes 2 and 5), and the 10.4-kbp pgk-2 gene inserted into pSP64 (lanes 3 and 6). Stable transformants of the NIH 3T3 cells were selected in G418, the cells were pooled, and extracts were prepared. An extract prepared from the testis of a strain C3H/He mouse was run in lane 8. Gel electrophoresis conditions and staining for PGK activity have been described (Bucher et al., 1980; Paterno and McBurney, 1985)). Only the portion of the gel containing PGK-2 is shown. The more slowly migrating PGK-1 human and mouse enzymes were approximately 10 times more active than the PGK-2 enzymes identified in lanes 1,3, and 6. The arrows indicate the locations of the PGK-2A and PGK-2B allelic variants. The expression of the pCASV2pgk-2 plasmid was observed in transiently (lane 1) but not stably (lanes 2 and 4) transfected cells, while expression of the stably integrated pgk-2 gene (lanes 3 and 6) was higher than in transiently transfected cells (data not shown).

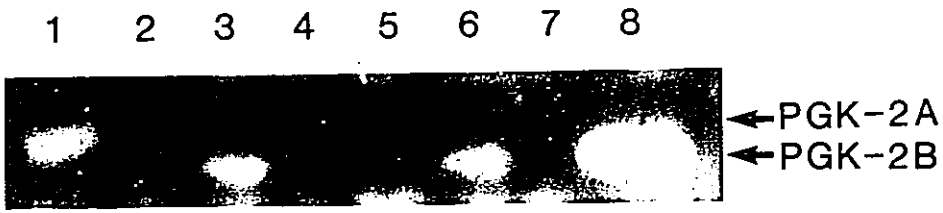

1 2 3 4 5 6 7 8
PGK-2A
PGK-2B

Fig. 5.11: The recombinant PCAB12-SP64 used in the transfection studies. The recombinant lambda genome carrying the B12 fragment was digested with BamHI then the insert was subcloned into the BamHI site of the plasmid pSP64 using the strategy described in section 2.8. Filled box represents hybridizing region; the striped boxes represent fragments which contained repeated sequences. B, BamHI; E, EcoRI, G, BglII; H, HindIII; p, PstI, X, XbaI.

pCA B12-SP64

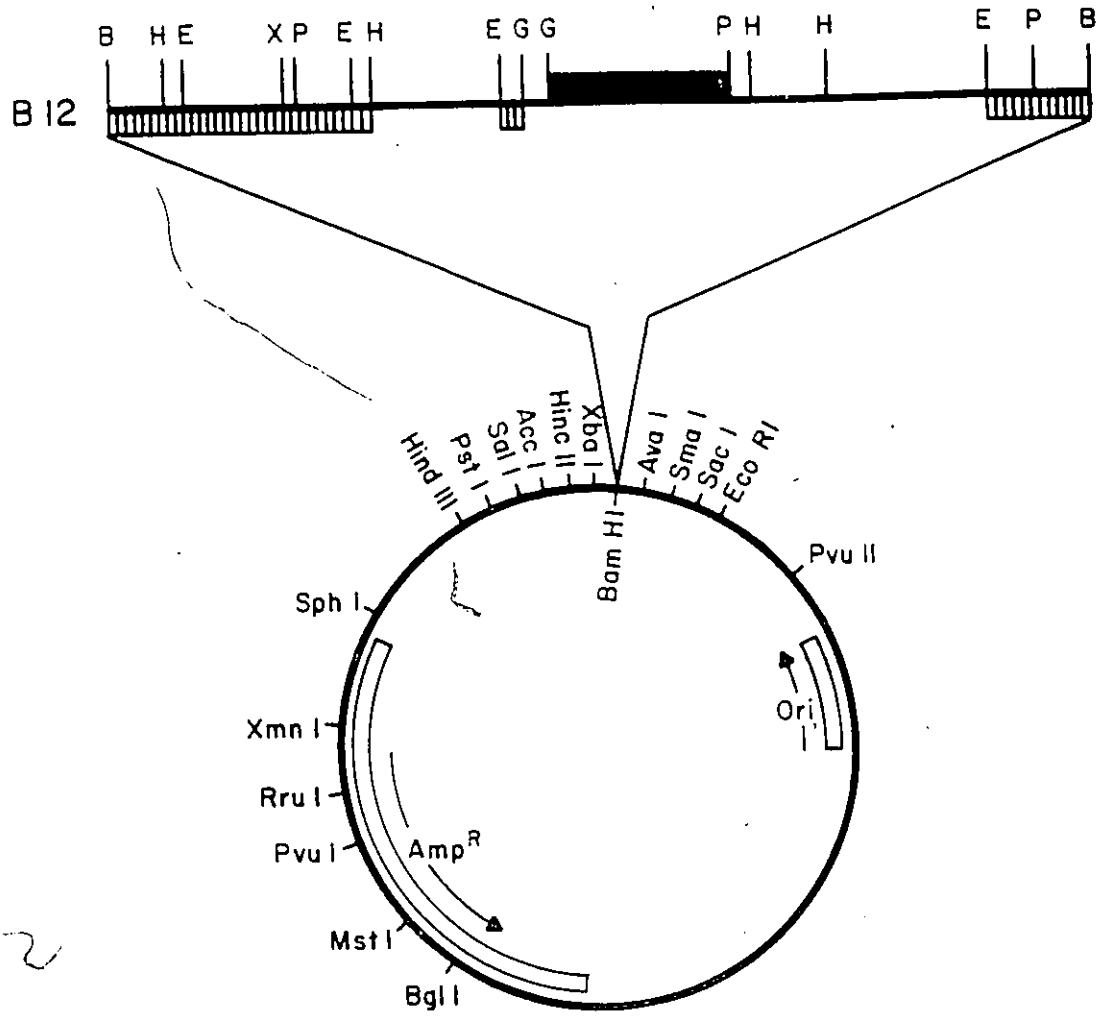
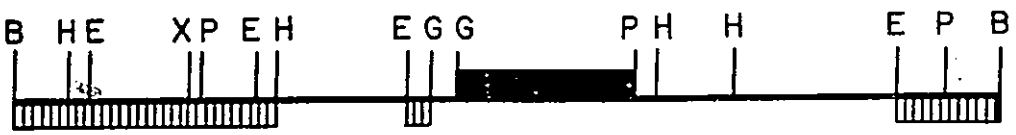
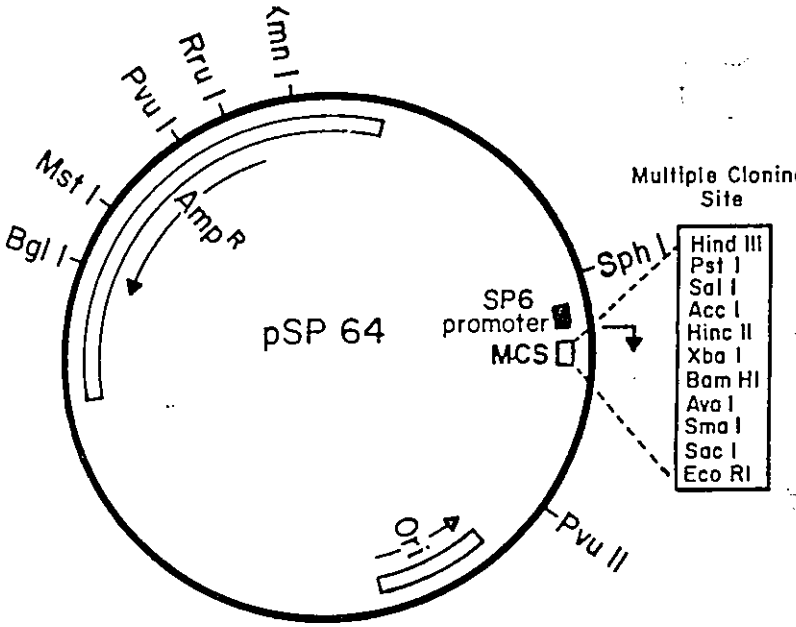

B12
B H E X P E H E G G P H H E P B
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Bam HI
Ava I
Sma I
Sac I
Eco RI
Pvu II
Ori
Sph I
Xmn I
Rru I
Pvu I
Amp^R
Mst I
Bgl I

Fig. 5.12: The recombinant pCAB12Bgl/P2.1-SP64 used in restriction map analysis. the B12 genomic fragment contained a short region of homology to the pgk-2 cDNA: 2.1-kbp PstI/BglII fragment. B12 was digested to completion with PstI and BglII. The 2.1-kbp fragment was ligated into a PstI/BglII digested pSP64. The 2.1-Kbp fragment is indicated as solid box; the fragments which contained repeated sequences are indicated by striped boxes. B, BamHI; E, EcoRI; G, BglII; H, HindIII; P, PstI; X, XbaI.

pCA B12-SP64 insert

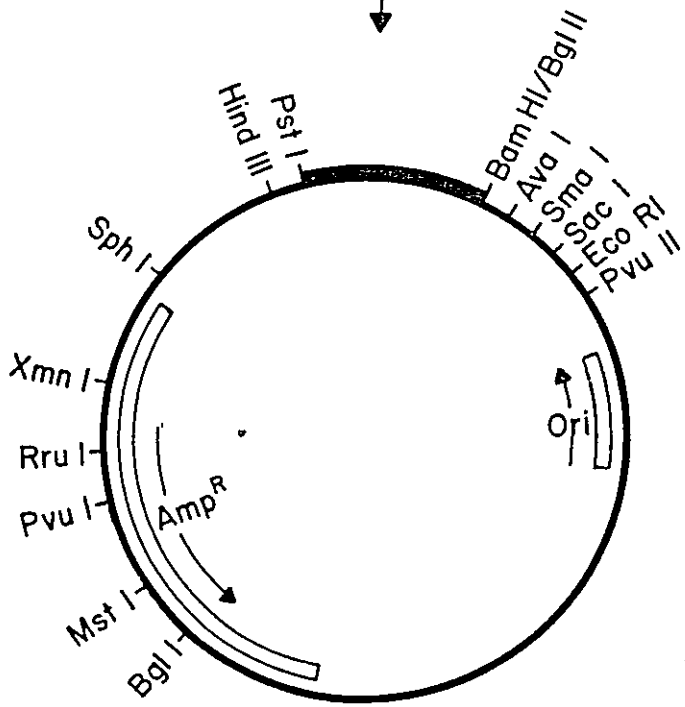




Bgl II + Pst I

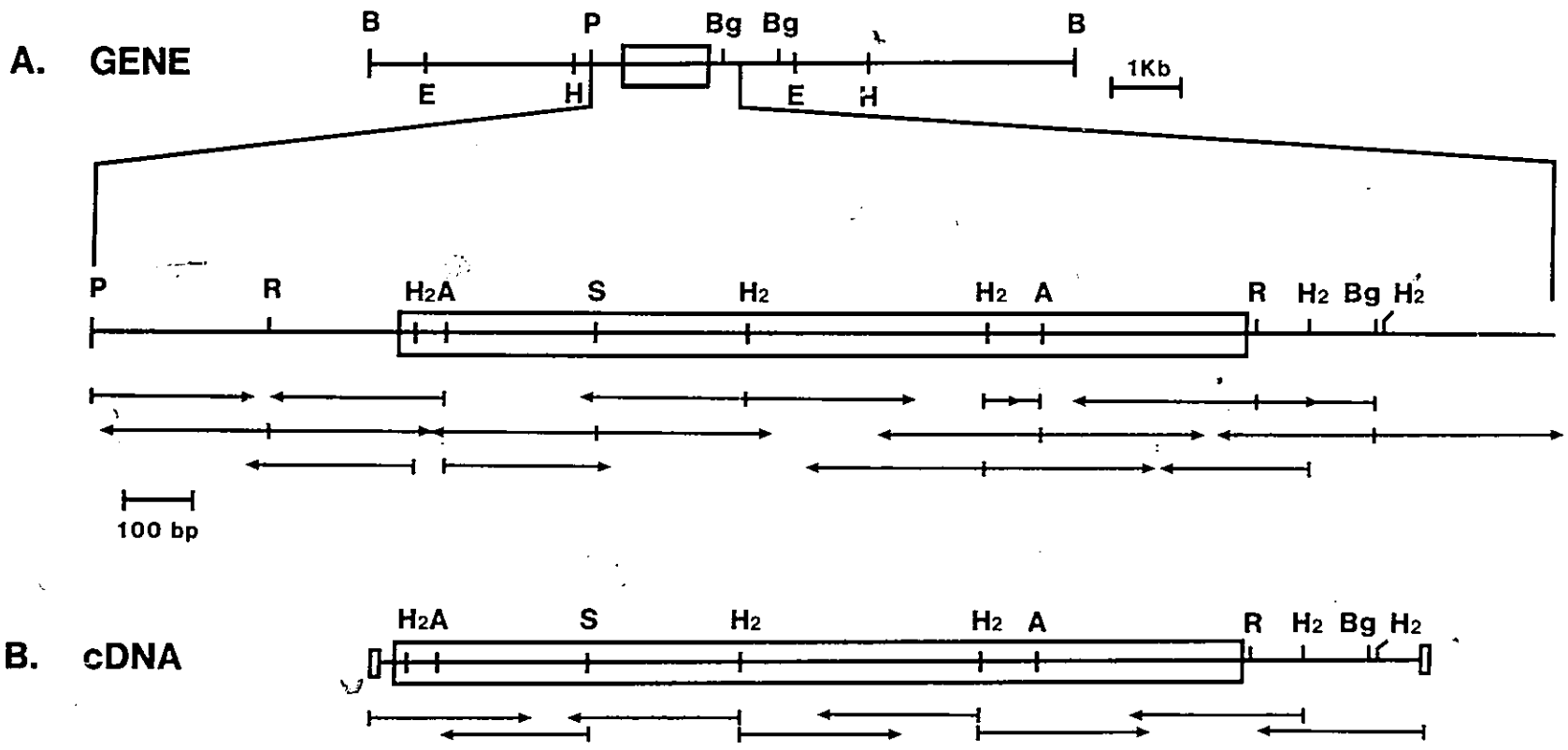
Bam HI + Pst I + Calf intestinal phosphatase

T4 DNA Ligase



pCA B12 Bgl/P 2.1-SP64

Fig. 5.13: Restriction maps of the pgk-2 gene and cDNA are colinear. The coding regions of the pgk-2 gene (A) and cDNA (B) are shown as open boxes; transcription is from left to right. Restriction sites: A, AccI; B, BamHI; Bg, BglII; E, EcoRI; H, HindIII; H_2, HincII; P, PstI; R, RsaI; S, SstI. Shown below each restriction map is the sequencing strategy. The arrows indicate the direction and extent of sequence determined.


A. GENE
B
P
Bg
Bg
B
E
H
E
H
1Kb
P
R
H2A
S
H2
H2
A
R
H2
Bg
H2
100 bp
B. cDNA
H2A
S
H2
H2
A
R
H2
Bg
H2

5.11 Nucleotide sequence of the pgk-2 cDNA and the pgk-2 gene

The gene and the pMPGK-2 insert were sequenced in collaboration with Dr. P. Boer according to the strategy outlined in Fig. 5.13. The sequence is shown in Fig. 5.14. The two sequences are identical except at position +452, at which the strain BALB/c codon is CAA (glutamine) and the strain C3H/He codon is CGA (arginine). The difference in the electrophoretic mobility between the two genetic variants of PGK-2 is the consequence of this mutation.

The amino acid composition of the mouse PGK-2 is highly similar to that of other PGKs (Mori et al., 1986). The coding region of the mouse pgk-2 gene and the pgk-1 cDNA (Mori et al., 1986) shows 80% nt identity (identical nucleotides are underlined in Fig. 5.14). The 3' noncoding region is 51% homologous (Fig. 5.15) whereas the 5' control region of pgk-2 shows no homology at all to that of mouse pgk-1 (Adra et al., 1987). The sequence of the mouse pgk-2 5' control region contains a number of sequence motifs. A CCAAT motif is present at -160 relative to the translation start site and a putative transcription start point is present at nt -57 (Corden et al., 1980). There is a GC box at nt -70 which represents a potential binding site for the Sp1 transcription factor (Dynan and Tjian, 1985). There is no TATA box.

Fig. 5.14: Nucleotide sequence of the mouse pgk-2 gene. The nucleotides are numbered from the start codon, and the deduced amino acids for PGK-2 are shown below. The entire sequence could be aligned with the human pgk sequence (Tani et al., 1985) to which we added G residues at positions 1021 and 1187. A dot over a nucleotide indicates identity with the human sequence, and small insertions of one to seven nucleotides in the human sequence are indicated by downward-pointing arrowheads. The boxed motifs are the common regulatory signals and include a CCAAT box at -160, a putative transcription start site at -57, and a polyadenylation signal at +1494. The poly(A) tail in the cDNA was added at position +1509 at the site of the large open arrow. The extent of the pgk-2 cDNA is indicated by large open arrows at positions -20 and +1509. A mouse strain polymorphism occurs at position +452 (indicated by a box), at which strain C3H/He codon is CGA and strain BALB/c codon is CAA. Between nucleotides -23 and +1640, the pgk-2 sequence could also be aligned with that of mouse pgk-1 cDNA (Mori et al., 1986). Here, identical nucleotides are underlined, and insertions of one to seven nucleotides in the cDNA sequence are indicated by upward-pointing arrow- heads. The asterisks at +1642 indicate the adenine-rich region, and the box at +1658 indicates the downstream direct repeat proposed by Tani et al. (1985). The broken-line box at -77 is a GC box which represents a potential binding site for the Sp1 transcription factor (Dynan and Tjian, 1985).

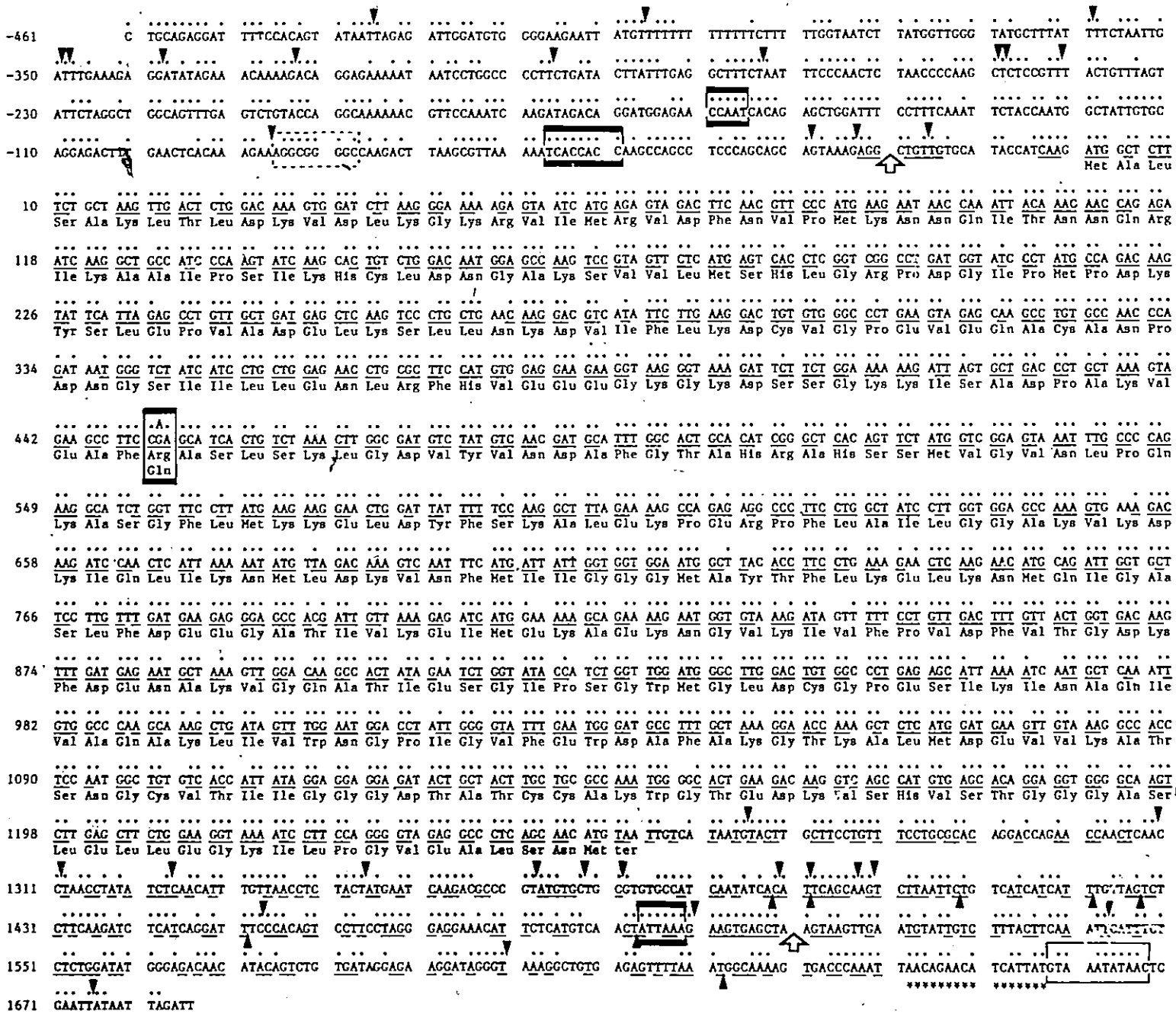
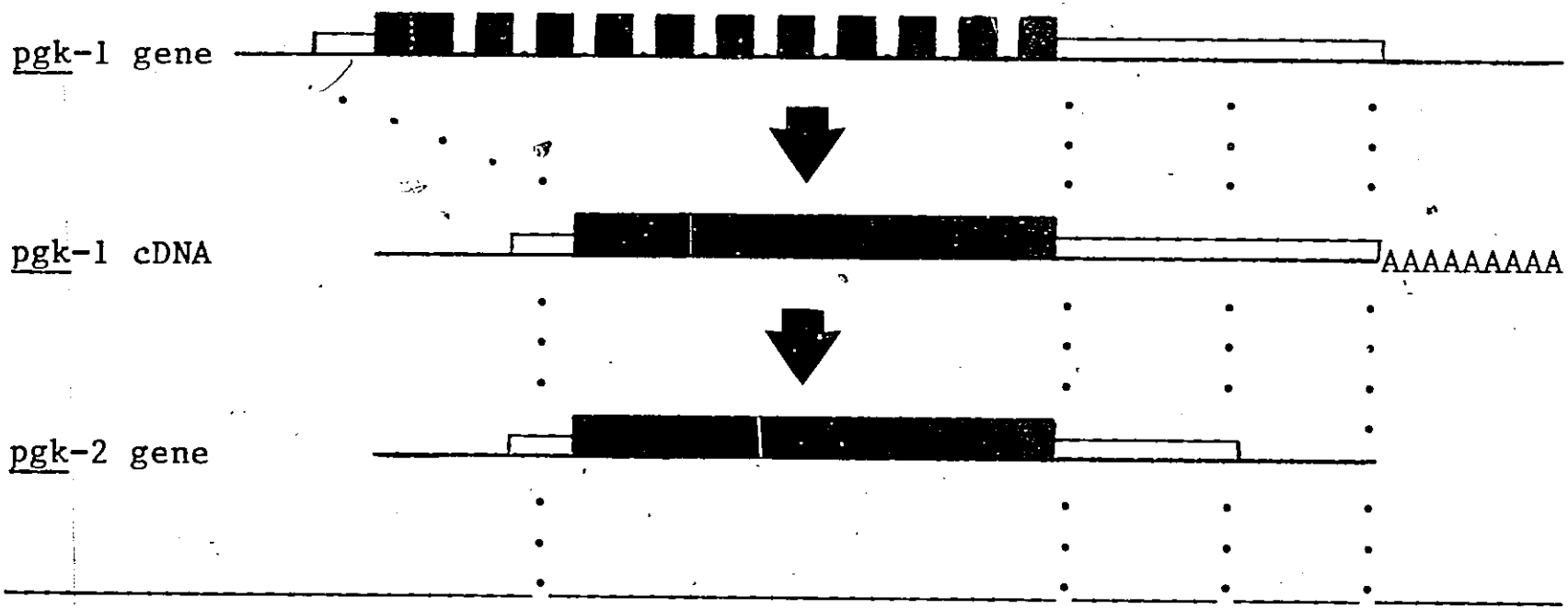


Fig. 5.15: Relationship between nucleotide sequences of pgk genes. (A) Model for the presumed origin of pgk-2. The ancestral pgk-1 gene, like the human and mouse genes, is presumed to contain 10 introns. The processed pgk-1 mRNA is thought to have been copied into a cDNA sequence which was integrated into an autosome. The new polyadenylation signal evolved some 130 nucleotides upstream of that used by pgk-1. Coding regions are shown as solid boxes, and nontranslated regions of mRNAs are indicated by open boxes. (B) Nucleotide (nt) sequence comparison over four regions of the pgk-2 gene, the upstream, coding, 3' noncoding, and downstream regions. The human pgk-2 sequence is from Tani et al. (1985) and the mouse pgk-1 cDNA sequence is from Mori et al. (1986). The mouse pgk-1 5' region (Chapter 3), like that of human pgk-1, is a G+C-rich sequence that shows no homology to the 5' region of pgk-2.

A



B

length (nt)	462	1251	259	130
pgk-2^m/pgk-1^m	none	80%	51%	47%
pgk-2^m/pgk-2^h	66%	85%	63%	51%

During the progress of my work on the PGK gene family a study was reported by Tani et al. (1985) in which they described the cloning and the characterization of a human processed pseudogene for PGK located on chromosome 6 near the histocompatibility genes. The validity of this conclusion was questionable. The mapping of B12 to the mouse chromosome 17, the strong hybridization of B12 to the pMPGK-2 cDNA insert, the fact that the 2.1.Kbp PstI-BglII fragment of B12 appears to be intronless, the linkage of the mouse pgk-2 gene to the major histocompatibility complex (Eicher et al., 1978), and my discussions of this work with Drs. J.S. Sam, A. Yoshida, N. Ellis, Y.-F. Lau and J. McCarry made me conclude that the sequence of the human clone contained errors responsible for the generation of an in-phase termination codon and that the human clone is in fact the human pgk-2 gene. This inference was verified by comparing the sequence of the so-called human processed pseudogene for PGK with the mouse pgk-2 sequence. The comparison revealed an 85% homology at the coding nucleotide level. A homology was also found at the 5'- and 3'-flanking regions (Fig. 15). The presumed functional motifs (Fig.14), the CCAAT box, the transcription start site and the polyadenylation signal (Birnstiel et al., 1985), are particularly conserved. These results suggested that the human pgk-related sequence actually codes for the testis-specific PGK. To confirm this suggestion, I cloned the human sequence in collaboration with Dr. R. Korneluk and this sequence was introduced by Dr. M.W. McBurney into EC cells. The transfected cells contained a PGK activity with mobility similar to that of the mouse PGK-2A isozyme.

DISCUSSION

5.12 The evolution of the pgk-2 gene

VandeBerg et al. (1973) discovered that there are two PGK isozymes, the PGK-1 and the PGK-2. The PGK-2 isozyme is usually expressed in spermatozoa of metatherians and eutherians and is specified by an autosomal gene. On the basis of the high level of homology, and the fact that the PGK-2 isoform is absent from prototherians, fish, birds and reptiles (VandeBerg, 1985), it can be concluded that the pgk-2 gene arose by a gene duplication before the divergence of metatherians from eutherians but after the divergence of birds and prototherians from the therians. Thus, this gene duplication event must have occurred between 100 (Cooper et al., 1977) and 200 (Whittaker and Thompson, 1974) million years ago (VandeBerg, 1985).

Further evidence for this conclusion is provided by sequence analyses. When we compared the sequence of the 3'-flanking region of the human and mouse pgk-2 genes using the analyses of Soares et al. (1985), we deduced that these two genes arose from a common gene and started diverging in the late Cretaceous (Romer, 1966) between 80 and 100 million years ago, at the time when the rodent and human orders diverged.

5.13 The pgk-2 gene is a recruited retroposon

The mouse pgk-2 gene is not interrupted by introns whereas the

human (Michelson et al., 1985) and the mouse (Chapter 3) pgk-1 genes contain 10 introns. On the basis of this observation and the following observations: a) the high degree of sequence identity between the mouse pgk-2 gene and the mouse pgk-1 cDNA over the coding region, b) the pgk-2 gene is located on chromosome 17, while the pgk-1 gene is X-linked, we may conclude that the pgk-2 gene is a transcribed retroposon derived from pgk-1. Some homology between the pgk-2 gene and the pgk-1 cDNA is found in the 3'-flanking region downstream of the pgk-2 polyadenylation signal (Fig. 5.14) indicated by a box. In addition the pgk-2 mRNA is about 130 nucleotides shorter than the pgk-1 mRNA (Fig. 5.8B and C). Thus, a new polyadenylation signal may have originated in the pgk-2 retroposon.

5.14 The pgk-2 gene lacks some of the hallmarks of a retroposon

The pgk-2 sequence does, however, lack some of the characteristics of retroposons (Rogers, 1985; Vanun, 1985). There is no poly(A) region at the 3' end of the pgk-2 gene. In addition, there are no flanking direct repeats. Because of the age of the pgk-2 gene, one would predict that the direct repeats and the A-rich sequence would have undergone significant divergence that they are no longer evident. However, an adenine-rich region located at nt +1642 (Fig. 5.14) could represent a remnant of the poly(A^+) tail.

5.15 The appearance of pgk-2 must have predated the evolution of X inactivation in spermatogenesis

A large number of retroposons have been cloned and characterized

(Vanin, 1985; Rogers, 1985). These sequences are generally not functional genes. Therefore one is forced to ask what mechanisms are responsible for the conservation of this rare functional retroposon? The mammalian X chromosome is subject to X inactivation during spermatogenesis, and all X-linked genes, including pgk-1, are turned off (Monesi, 1971). Mammalian sperm cells rely on glycolysis mediated by PGK-2 for ATP generation (VandeBerg, 1985). Therefore the inactivation of the X chromosome in spermatogenetic cells must have provided the selective pressure to maintain the pgk-2 retroposon active, thereby enabling sperm cells to metabolize fructose.

Our observations suggest that retroposition gives rise occasionally to functional genes and that the appearance of X chromosome inactivation postdated the evolution of the pgk-2 gene.

5.16 B6, B8, B9, B10, B13, B15 and B24 are likely to be pgk-1 retroposons:

The genomes of Balb/c and C3H/He strain mice appear to contain eight regions or loci homologous to the pgk-1 cDNA probes. I have discussed previously the cloning and the characterization of both the pgk-1 and the pgk-2 genes.

The other regions of the mouse genome homologous to the pgk-1 cDNA probes are likely to be retroposons derived more recently from the pgk-1 gene. These pgk-1 related sequences are dispersed onto at least three different autosomal chromosomes. Studies of band intensities in male and female mouse DNAs indicated that none are

X-linked like the parental pgk-1 gene (Chapter 5, section 5.3). All of these presumptive pseudogenes contain the pgk-1-related sequence within a relatively short stretch of DNA and in no case is this pgk-1 related sequence interspersed with repeated sequence elements. These results suggest that none of the pgk-related sequences is interrupted by introns.

5.17 B13 and B15 seem to be recent classical processed pseudogenes:

The two fragments B13 and B15 are most likely to contain classical processed pseudogenes. The B13 fragment hybridized to the pgk-1 cDNA probes even under very high stringency conditions, indicating a high degree of sequence homology. In addition, the B13 and B15 sequences hybridized to probes specific for all regions of the pgk-1 cDNA. Finally, the region of homology to the pgk-1 cDNAs was very short, in the case of B13 less than 1.5 kb and for B15 between 1.4 and 3.6 kb. The recent evolution of B15 is also consistent with the fact that this fragment is found in Balb/c but not C3H/He strain mice, indicating that it has not become fixed in the Mus musculus population. The C3H/He-specific fragment, B2, is not allelic to B15 and must also have arisen recently. We have relatively little information concerning the B2 fragment because it was not cloned.

5.18 B6, B8, B10 and B24 are truncated pgk-1 pseudogenes:

The other four cloned fragments, B6, B8, B10, and B24, carried truncated copies of the pgk-related sequence since each of these

regions failed to hybridize to probes specific to either the 5', the 3', or both ends of the cDNA. Both B6 and B24 contained sequences highly homologous to the pgk-1 cDNA probe, indicating a relatively recent origin.

5.19 B24 is an unusual retroposed pseudogene:

The sequence of B24 argues strongly that the pgk-related region is indeed a retroposed pseudogene derived from pgk-1. The degree of sequence identity is very high, but termination codons in all reading frames indicate that a protein could not be encoded by this region. The B24 region is located on chromosome 16, while the pgk-1 gene is X-linked and the B24 sequence is not interrupted by introns. This B24 sequence does, however, lack some of the hallmarks of retroposed pseudogenes. The B24 sequence is truncated at its 3' end, an unusual but not unique characteristic of processed pseudogenes (Vanin, 1985; Nishioka et al, 1980; Ullu and Weiner, 1984). This 3' truncation means there is no A-rich sequence at the 3' end of the pgk-related region. In addition, there are no direct repeats flanking the pgk-related sequence.

Given the high degree of sequence homology between the pgk-1 cDNA and B24, which suggests relatively recent evolution of this latter sequence, one would have expected the direct repeats to be evident. Although other retroposons lacking direct repeats have been reported (Denison and Weiner, 1982), there are some unusual sequences downstream of the pgk-related sequence in B24 consistent with the idea that the B24 fragment might come from an area of the genome active in

recombination. Thus, the B24 region may have originated as a nontruncated pgk-1 retroposon from which the 3' and possibly 5' ends were subsequently lost by recombination. Regions of the genome particularly active in recombination have been inferred from the clustering of retroposons (Rogers, 1985).

5.20 Retroposition occurs in the female or embryonic germ lines but not the postmitotic male germ line:

Retroposons are present in the germ-line DNA and are thought to have originated by reverse transcription of mRNA species in germ-line cells. The pgk-2 gene is expressed exclusively in male germ cells. When mouse genomic DNA was probed with the full length pgk-2 cDNA (probe G in Fig. 1), only the B12 and B17 fragments hybridized, and when the same blots were probed with a region comprising the 3'-untranslated region of the pgk-2 mRNA, only the B12 fragment hybridized (Boer et al., 1987), indicating the absence of pgk-2 derived pseudogenes. The abundance of pgk-1 pseudogenes along with the absence of pgk-2 pseudogenes may indicate that conditions for successful retroposition are not present in those stages of spermatogenesis during which the pgk-2 gene is normally expressed. The presence of pseudogenes from the mouse somatic cytochrome c gene but not from the sperm-specific cytochrome c gene (Limbach and Wu, 1985) is also consistent with this notion.

5.21 Mechanisms at work for removal of unnecessary sequences:

The human genome contains only four sequences related to the

pgk-1 cDNA (Michelson et al., 1985). The human and mouse pgk-1 genes are X-linked and the human and mouse pgk-2 genes are both linked to the major histocompatibility complex. The two human pseudogenes are on chromosomes X and 19 (Willard et al., 1985; R.G. Korneluk, C.N. Adra and A.G. Hunter, manuscript in preparation). No mouse pseudogene is located on the X chromosome, and the mouse genome contains several more pseudogenes than does the human genome. These species differences had been noted for other gene families (Piechaczyk et al., 1984) and are consistent with the notion that most retroposed pseudogenes have arisen relatively recently in evolution, that is since the divergence of most of the present mammalian species. The exception to this is the pgk-2 gene which arose by retroposition more than 100 million years ago. Given that retroposition has been possible for at least 100 million years, it seems curious that most pseudogenes appear to have evolved so recently (Vanin, 1985). The trivial explanation for this might be that very old retroposed pseudogenes have diverged so much from the parental gene that their reactivity with cDNA probes is too low to be detected. An alternative explanation might be that these nonfunctional pseudogenes are removed from the genome. Consistent with this latter notion are our observations on the B24 sequence, which suggest that the 3' and possibly 5' regions of the original retroposed pseudogene may have been removed following its integration into the genome.

The family of pgk-1-related sequences in the mouse appears to be a record of retroposed sequences originating from the pgk-1 gene at various times during the past 100 million years. Our results suggest that the genome is evolving continuously with the formation and perhaps removal of retroposed sequence elements.

CHAPTER VI

THE FUTURE

6.1 A model for the control of X chromosome activity

From our studies described in the previous chapters (also see Appendix 1) and from various studies demonstrating that the transcription of genes is influenced by the binding of regulatory proteins to some specific DNA sequences (Ptashne, 1986; Echols, 1986; Schleif, 1987), it is possible to propose the following model as a molecular explanation for X chromosome inactivation:

It is proposed that an autosomal gene produces an inactivation factor (perhaps several autosomal genes may interact with X-linked genes to produce the inactivation factor(s)) that has the ability to interact with specific DNA sites present at the Xic and near the promoter of each X-linked gene. The formation of DNA loops and the cooperativity (Schleif, 1987) inherent in such a system could allow the transcriptional regulation of the X chromosome.

Clearly, the availability of the mouse pgk-1 gene in combination with the EC cell lines should now permit the verification of these inferences.

Perspectives and experimental approaches

6.2 Fitting methylation into X chromosome inactivation:

A number of experiments should be done before we can develop a

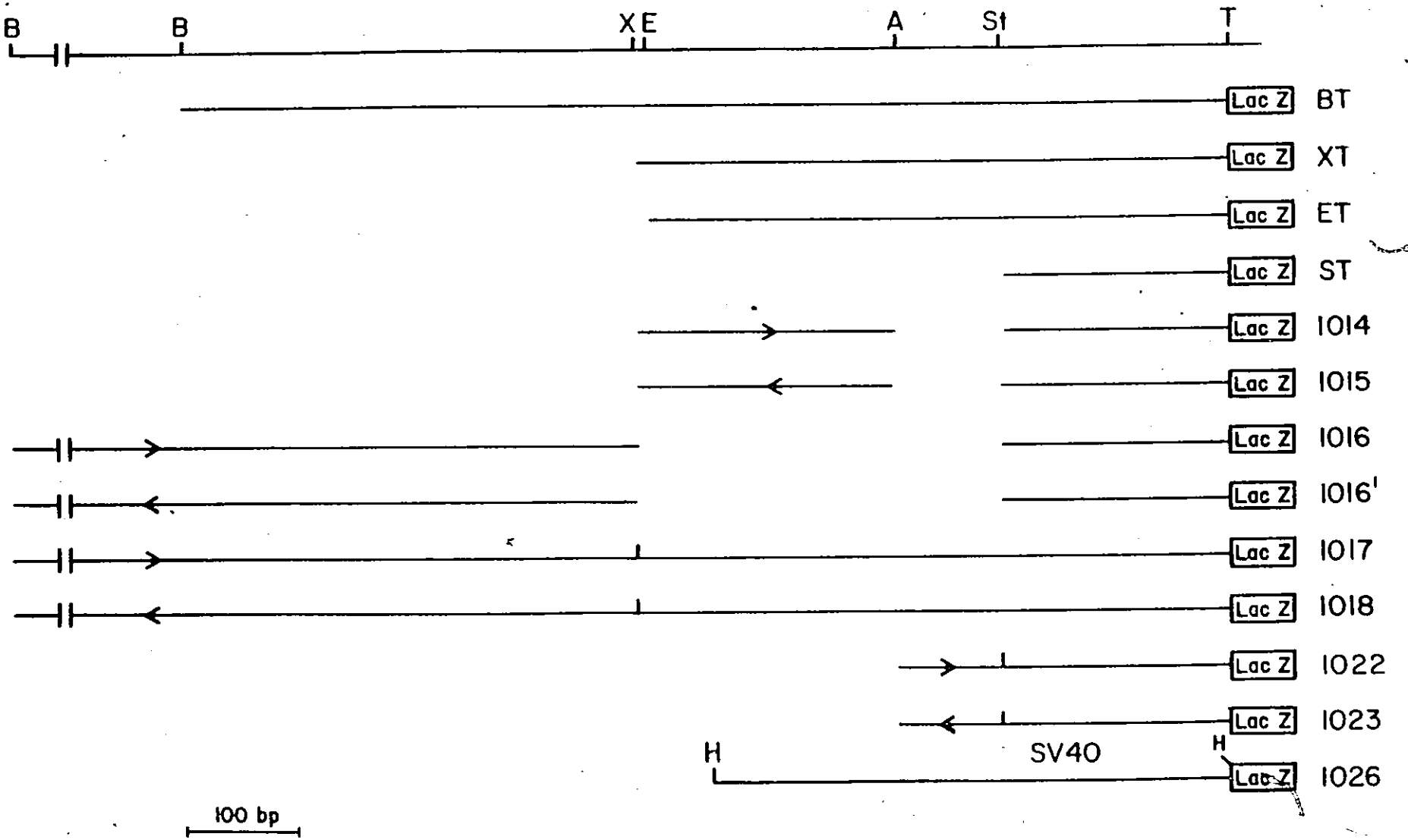
clear picture of the role of methylation in X chromosome inactivation and in turning on and off genes. My analysis of the mouse pgk-1 upstream region indicated that the CCGG and GCGC sites assayed for methylation are unmethylated on both the active and inactive X chromosomes. However, it is possible that these sites are not critical for the process of dosage compensation. Therefore it is important to look for methylation differences around the pgk-1 gene in our EC cells and in the developing female mouse embryos. Using pulsed-field gradient gel electrophoresis (Schwartz and Cantor, 1984; Carle et al., 1986; Smith) and single copy probes that I have derived from the mouse pgk-1 gene (data not shown), Mrs. Molly Bartlett is now determining the position of C+G-rich DNA (Brown and Bird, 1986) and its methylation pattern before, during and after X chromosome inactivation.

6.3 Analysis of the 5' control region of the mouse pgk-1 gene:

I have initiated the functional characterization of the promoter region with the assumption that it will allow us to address questions concerning not only the regulation of the expression of the pgk-1 gene, but also the mechanism of X chromosome differentiation.

I have constructed 12 chimeric genes (Appendix 2) by fusing various fragments of 5'-flanking sequences to the *E coli* Lac Z gene (Fig. 6.1). Lac Z encodes beta-galactosidase. These constructs can be tested for expression *in vivo* by *in situ* staining or by spectrophotometric assays. All these constructs were transfected into mouse P19 embryonal carcinoma cells. Preliminary results obtained by I and Dr. M.W. McBurney suggest that all constructs are functional in transient

Fig. 6.1: The five constructs 1017, BT, XT, ET and ST contained 5.4 kb, 854, 455, 435 and 135 bp of sequences upstream of the Taq I site. Construction of 1014 and 1015 was done by inserting the XbaI/AluI fragment in the natural and in the reverse orientations respectively in front of the ST recombinant. 1016 and 1016' were constructed by fusing the 4.5 kb BamHI/XbaI fragment in the natural and in the reverse orientations respectively into the ST construct. 1018 is identical to 1017 but contains the 4.5 Kb BamHI/XbaI fragment in the reverse orientation. The AluI/StuI fragment was ligated to the ST chimeric gene to yield two recombinants (1022, 1023) containing this sequence in either orientation. As a positive control in the transfection studies and to measure the activity of these various constructs I prepared the recombinant 1026 (Appendix 2) in which the Lac Z gene is under the control of the simian virus 40 promoter and enhancers.


B
B
X
E
A
St
T
Lac Z
BT
XT
ET
ST
1014
1015
1016
1016'
1017
1018
1022
1023
H
SV40
H
1026
100 bp

assays. In addition, our expression studies indicated the presence of a strong upstream activator element in the XbaI/StuI fragment. The ST construct which contains the CAAT motif and four of the five GC boxes has the same activity as the construct containing the SV40 promoter and enhancer (1026). When the largest construct (1017) was also introduced into P19 cells, its activity was approximately 30-fold higher than that of the SV40 construct (1026, see also Appendix 2). I have also made additional constructs to see if the XbaI/StuI fragment fulfill the strict criteria of an enhancer. They are currently being tested.

From these chimeric genes it will be possible to generate additional constructs, by deletion and insertion, to determine the identity of sequences that may be important for the regulation of the expression of the pgk-1 gene.

Our female EC cell lines represent different stages of X chromosome differentiation (Chapter I). If transcription repression is primarily determined by the interaction of factors with the X specific sequences one might observe, after transfection of the various recombinants into these cells, a pattern of activation that is related to the state of activity of the X chromosome. Moreover the study of the pgk-LacZ series may reveal which motifs are likely to be recognized by proteins.

6.4 Protein-DNA interaction studies:

In both prokaryotes and eukaryotes transcription seems to be influenced by proteins binding at a specific DNA site (Dynan and

Tjian, 1985; Ptashne, 1986; Richmond, 1987). As described previously, the transfection studies are already revealing the presence of upstream sequences regulating the pgk-1 expression. Proteins that bind to these cis-acting DNA sequences can be identified and eventually purified. I have subcloned short fragments (see Appendix 2) of the upstream region of pgk-1, in the vectors Gemini 3 and 4, to be used in the identification, purification and characterization of these proteins. By using gel retardation assays (Burke et al., 1987) on nuclear extracts prepared from retinoic acid-treated and untreated EC cells, it will be possible to detect DNA-binding proteins which interact with these subcloned fragments. Preliminary results obtained by Dr. P. Boer suggest the presence of factors that interact specifically with upstream elements. In addition, these studies will determine whether binding of these factors in vitro will correlate with transcriptional regulation of the transfected constructs. Nuclear extracts prepared from the four EC cell lines may not contain the same amount of pgk-1-binding activity relative to total nuclear protein. Purification of the identified regulatory proteins could be achieved by constructing a sequence-specific DNA affinity column (Kadonaga and Tjian, 1986).

6.5 Identification of X-linked sequences required for X chromosome inactivation:

The ability of X-linked gene to undergo X chromosome inactivation may be due to the presence of cis-acting regulatory sequences. Therefore, it will be of interest to introduce the pgk-LacZ constructs described in section 6.3 in P10 cells, to select clones with X-linked pgk-LacZ, and to study the behaviour of the chimeric gene on the inactive X chromosome. If differentiation leads to the inactivation

of pgk-LacZ chimeric gene then we can conclude that this gene has an element important in the control of its expression. Deletion analysis will be necessary to identify the sequences required for inactivation.

6.6 Isolation of the X inactivation center:

The central problem in X chromosome inactivation is to understand the event of the initiation of this process. Studies on X-autosome translocations in mice has mapped the Xic to the same site as the Xce. It appears that these two loci are identical. The Xic has been located near the pgk-1 locus.

There are techniques, now available, that will allow the detailed mapping of the mouse X-chromosome pgk-1 region. The combination of these techniques with cloning methods (Poustka et al., 1987) will have great potential for the isolation of long stretches of the mouse X chromosome. I was capable of cloning 45 Kb fragments in pSP64. I am currently trying the circularization of 100 Kb DNA fragments for the transformation of DH5 cells. Also, it is possible to construct linear DNA molecules that can be maintained in yeast as linear artificial chromosomes (Burke et al., 1987). Thus the Xic may be identified by fusing the cloned DNA segment to the pgk-1^b gene and introducing the vector into P10 cells carrying the pgk-1^a gene. Xce presence on the vector could be revealed by measuring the level of expression of pgk-1^b before and after differentiation. Our preliminary results do suggest that the Xic is not present in the pgk-1^b allele. Once isolated, the Xic will be extremely useful in DNA-binding protein studies. These studies will be the crucial test for the ideas discussed above.

APPENDIX 1

X chromosome reactivation in mouse embryonal carcinoma cells

The embryonal carcinoma cell line C86S1 has two X chromosomes, one of which is genetically inactive. C86S1A1(XX) is an hypoxanthine phosphoribosyl transferase (HPRT) deficient mutant which carries the $hprt^-$ allele on the active X chromosome and the $hprt^+$ allele on the inactive X chromosome. In these cells, DNA demethylating agents such as 5-azacytidine (5AC) induce the transient expression of the $hprt^+$ allele present on the inactive X chromosome in almost all cells and the stable expression of HPRT in up to 10% of surviving cells (Paterno et al., 1985). The RNA from stable clones of C86S1A1(XX) were analyzed by Northern blotting and probing with cDNAs to three X-linked genes: hprt, phosphoglycerate kinase (pgk-1) and ornithine carbamoyl transferase (oct). All stable clones had elevated levels of HPRT (Fig. A.1) and PGK-1 mRNA (Fig. A.2) consistent with their transcription from both X chromosomes, but the liver-specific X-linked gene oct (Fig. A.3) was not transcribed indicating that only appropriate X-linked genes were activated on the reactivated X chromosome. These results suggest that the 5AC treatment induced expression of genes on the reactivated X chromosome and that DNA methylation is involved in X chromosome inactivation. No such reactivation occurred in C86S1A1 cells if they had differentiated prior to treatment with 5AC. Thus, the process of X chromosome inactivation may be a sequential one involving, as a first step, methylation of certain DNA sequences and, as a second step, some other mechanism of transcriptional repression.

Fig. A.1: Level of hprt RNA in various EC cell lines and in HAT^R clones of C86S1A1. Total RNA was purified, fractionated on a 1% formaldehyde gel, blotted on nylon membrane, then hybridized to pHPT5 (Konecki et al., 1982). 50 µg were loaded in each lane. Lanes 1, C86S1; 2, $C86S1A1_2$(XO); 3, C86S1A1; 4, spontaneous HAT^R clone; 5, HAT^R clone AC1; 6 and 8, HAT^R clone 0A; 7 and 9, HAT^R clone 1B; 10, RJK88p18 (Chinese hamster cell line with a deleted hprt, Fuscoe et al., 1983). The filter was washed at 50°C in 1 x SSC + 0.1% SDS for 1 hour.

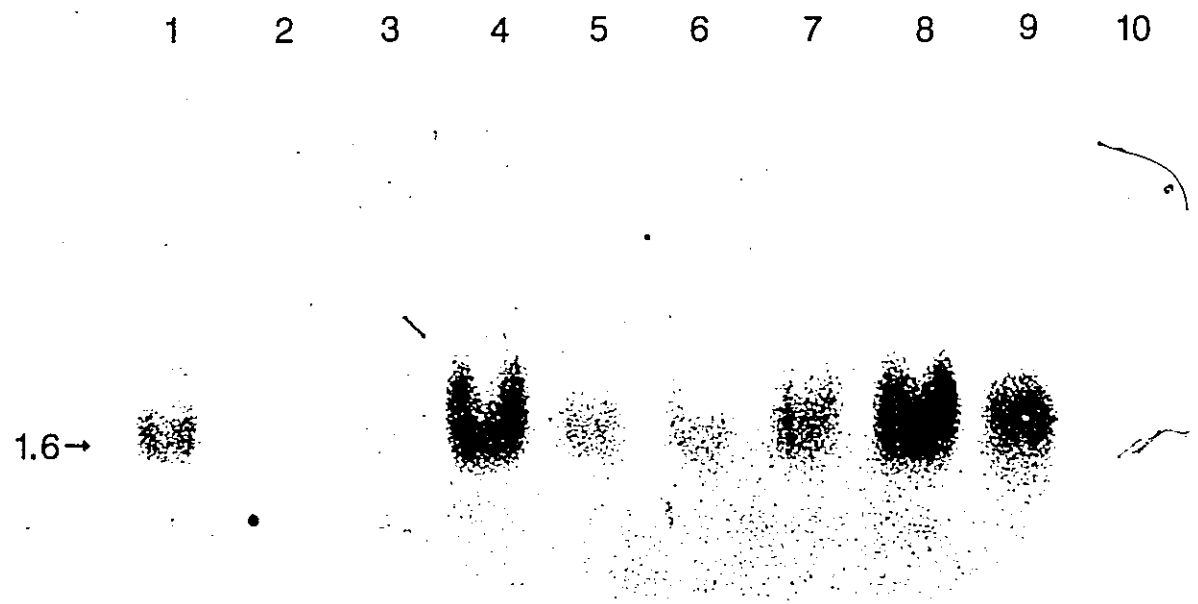
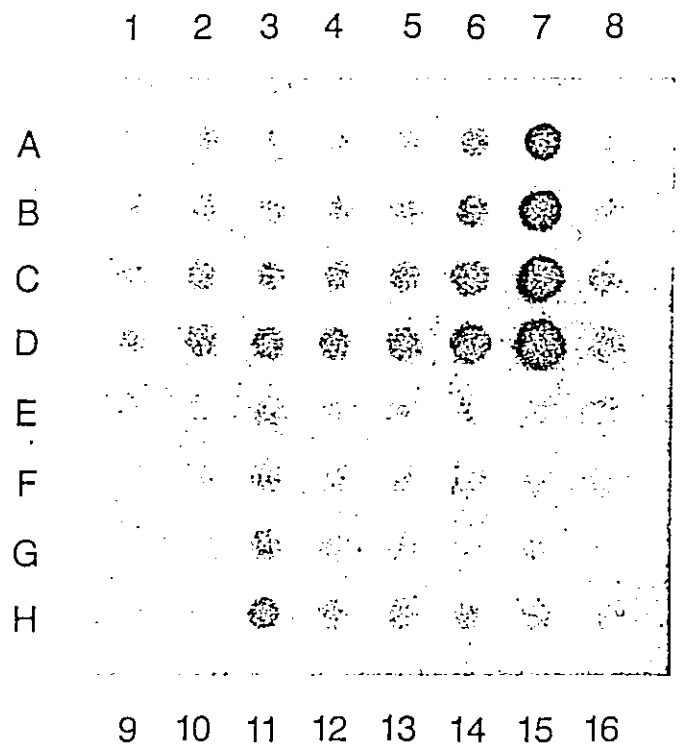

1
2
3
4
5
6
7
8
9
10
1.6→

Fig. A.2: Level of pgk-1 RNA in various EC cell lines and in reactivants of C86S1A1. Total RNA was purified, fractionated on a 1% formaldehyde gel, blotted on nylon membrane, then hybridized to the mouse pgk-1 cDNA insert (Adra et al., 1987). 50 µg were loaded in each lane. Lanes, 1, C86S1; 2, C86S1A1$_2$(XO); 3, C86S1A1; 4, spontaneous HAT^R clone; 5, HAT^R clone AC1; 6 and 8, HAT^R clone OA; 7 and 9, HAT^R clone 1B; 10, RJK88p18. The blot was washed at 50°C in 1 x SSC + 0.1% SDS.


1
2
3
4
5
6
7
8
9
10
1.9→

Fig. A.3: Level of oct RNA in various EC cell lines before and after exposure to 5-AC. Total RNA from EC cell lines was isolated, dotted onto nylon membrane, then hybridized at 42°C in the presence of 50% formamide with OCT nick-translated cDNA. Lanes, 1 and 9, C86S1; 2 and 13, $C86S1A1_2(XO)$ and $C86S1A1_3(XO)$ respectively; 3 and 11, C86S1A1; 4 and 5, C86S1A1AZAOA and C86S1A1AZA1B respectively; 6, RJK88p18; 7, liver; 8, spleen; 10 and 12, RNA purified from C86S1A1 cells which were treated with 5-AC for 24 hours; $C86S1A1B_2(XO)$, RNA purified from cells which were treated with 5-AC for 24 hours; 15, TG1a; 16, RNA purified from TG1a cells which were treated with 5-AC for 24 hours. Rows: A and E, 1.25 µg; B and F, 1.25 µg; C and G, 5 µg; D and H, 10 µg. The blot was washed at room temperature in 1 x SSC + 0.1% SDS for 1 hour.

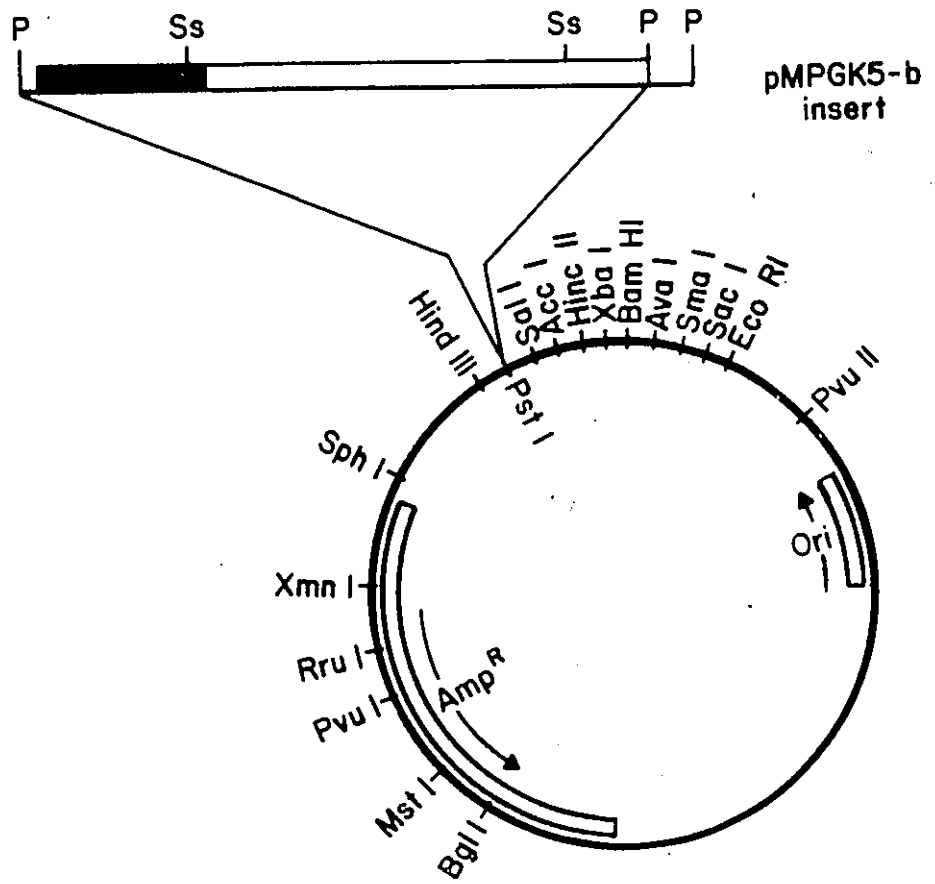

1 2 3 4 5 6 7 8
A
B
C
D
E
F
G
H
9 10 11 12 13 14 15 16

APPENDIX 2

List of Plasmids

The recombinant pCAM-pgk-1-3'-SP64 contained the 3' end of the mouse pgk-1 cDNA. This cDNA was sequenced in collaboration with Dr. P. Boer. Its sequence is identical to nucleotides 918-1639 of the cDNA reported by Mori et al. (1986). This construct was prepared by ligating the PstI fragment derived from pMPGK5-b (Adra et al., 1988) into a PstI-digested pSP64. Open box represents untranslated region; the filled box represents coding region. P, PstI; Ss, Sst I.

pCA M-pgK-1-3'-SP64

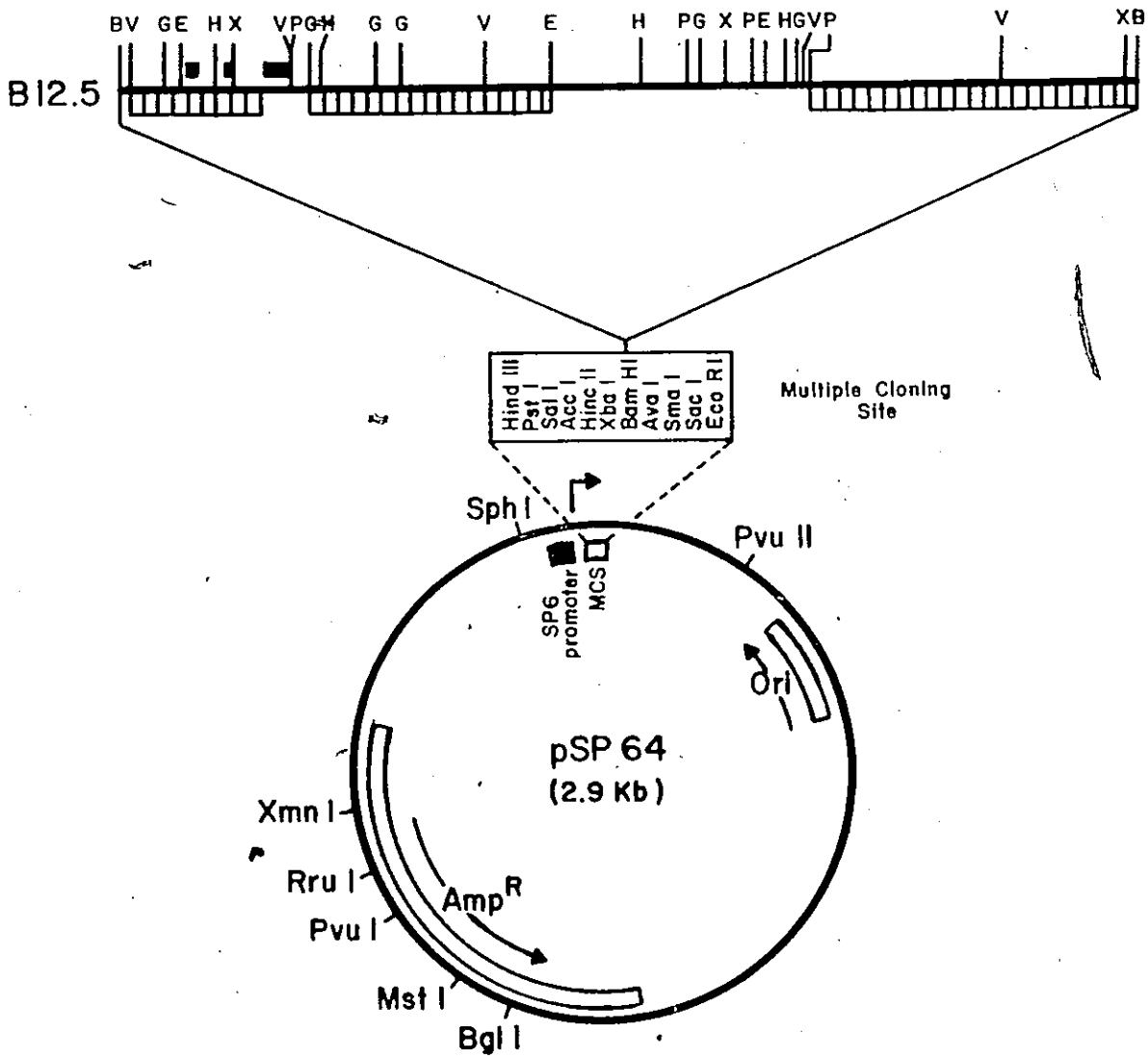


APPENDIX 2 Cont'd

List of plasmids

The recombinant pCA2P-64. It contains exons 9, 10 and 11 of the mouse pgk-1^b allele and approximately 7 Kbp of 3'-flanking sequences (Adra et al., 1987). The recombinant lambda genome carrying the B12.5 fragment was digested with BamHI then the insert was subcloned into the BamHI site of the plasmid pSP64 using the strategy described in section 2.8. Filled box represents hybridizing region; the striped boxes represent restriction fragments which contained repeated sequences. B, BamHI; E, EcoRI; G, BglII; H, HindIII; P, PstI; V, PvuII; X, XbaI.

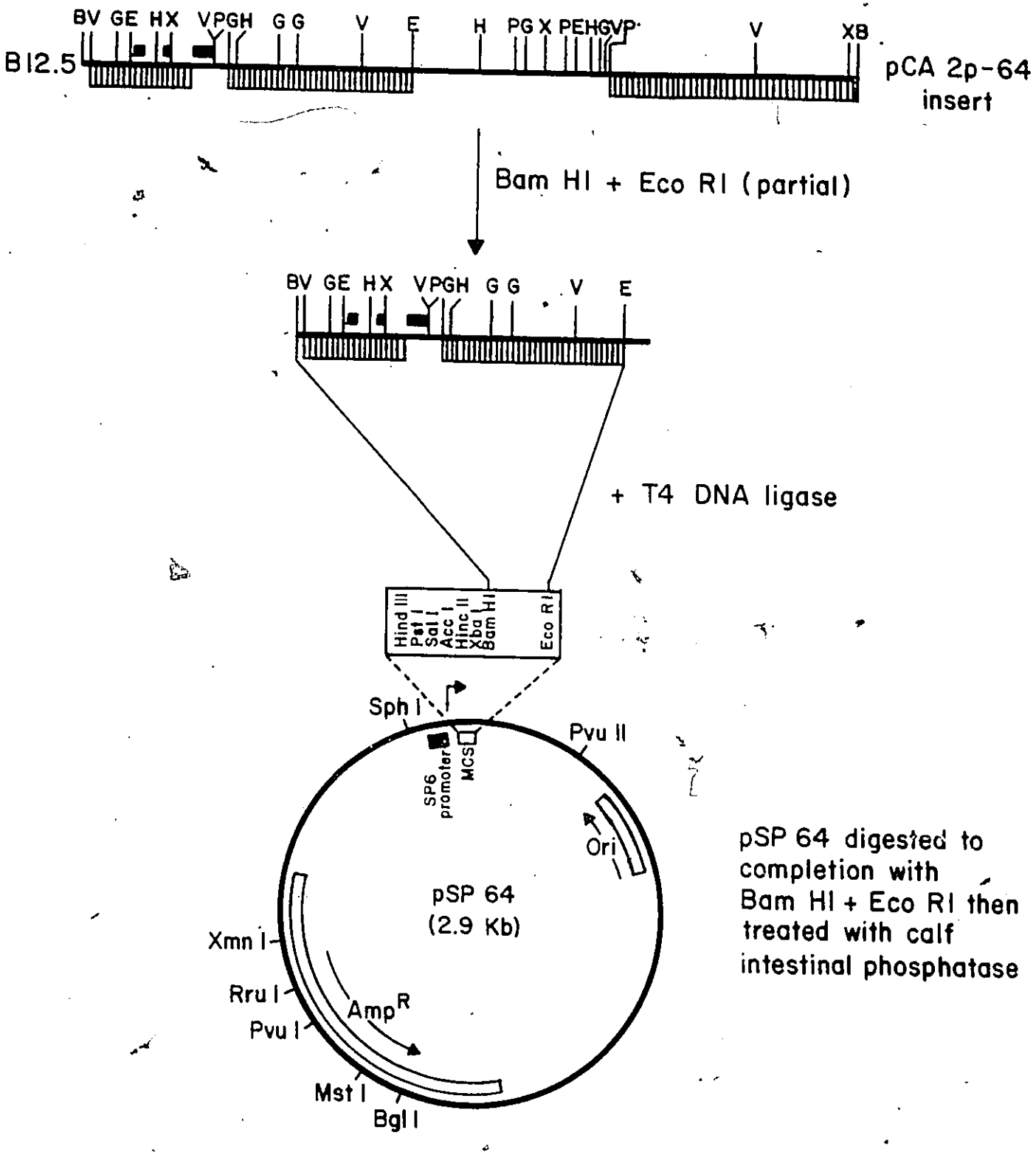
pCA2p-64


B12.5
BV GE HX VPGH G G V E H PG X PE HGVP V XB
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Bam HI
Ava I
Sma I
Sac I
Eco RI
Multiple Cloning Site
Sph I
Pvu II
SP6 promoter
MCS
Ori
pSP 64
(2.9 Kb)
Xmn I
Rru I
Pvu I
Mst I
Bgl I
Amp^R

APPENDIX 2 Cont'd

The recombinant pCAB12.5BamHI-EcoRI4.5-64. This construct contained the last three exons of the pgk-1^b allele and approximately 3Kb of 3'-flanking sequence. It was very useful in the reconstruction of the pgk-1^a and the pgk-1^b genes (Adra et al., 1987; also see Appendix 6). It was constructed by digesting pCA2p-64 (Appendix 3) to completion with BamHI and partially with EcoRI. The 4.5-Kb BamHI-EcoRI was excised from low melting agarose gel and ligated to BamHI-EcoRI-cleaved pSP64. Filled box represents hybridizing region; the striped boxes represent restriction fragments which contained repeated sequences (Adra et al., 1988). B, BamHI; E, EcoRI; G, BglII; H, HindIII; P, PstI; V, PvuII, X,XbaI.

pCA Bl2.5 Bam Hl-Eco Rl 4.5-64

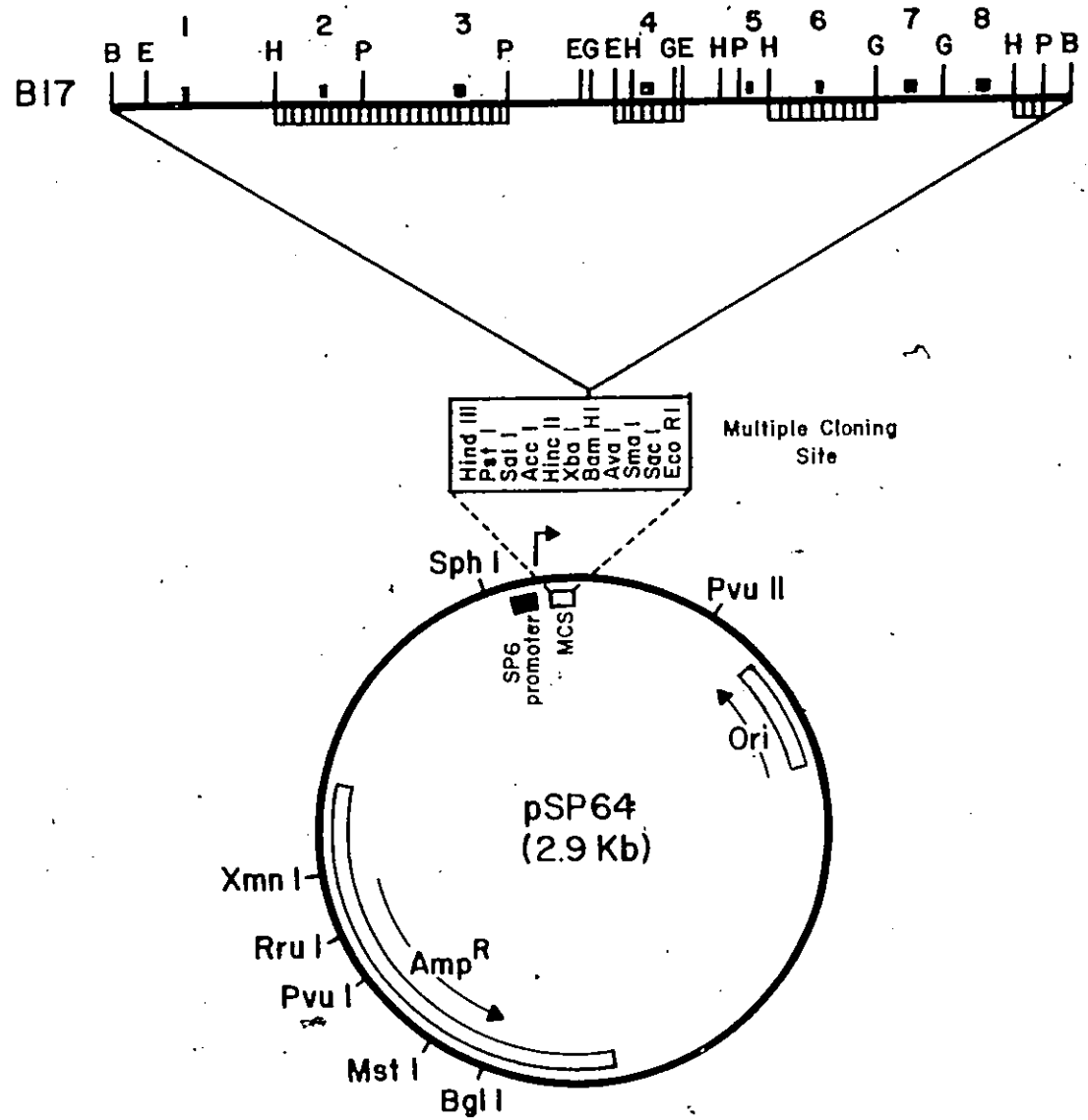


pSP 64 digested to completion with Bam Hl + Eco Rl then treated with calf intestinal phosphatase

APPENDIX 2 Cont'd

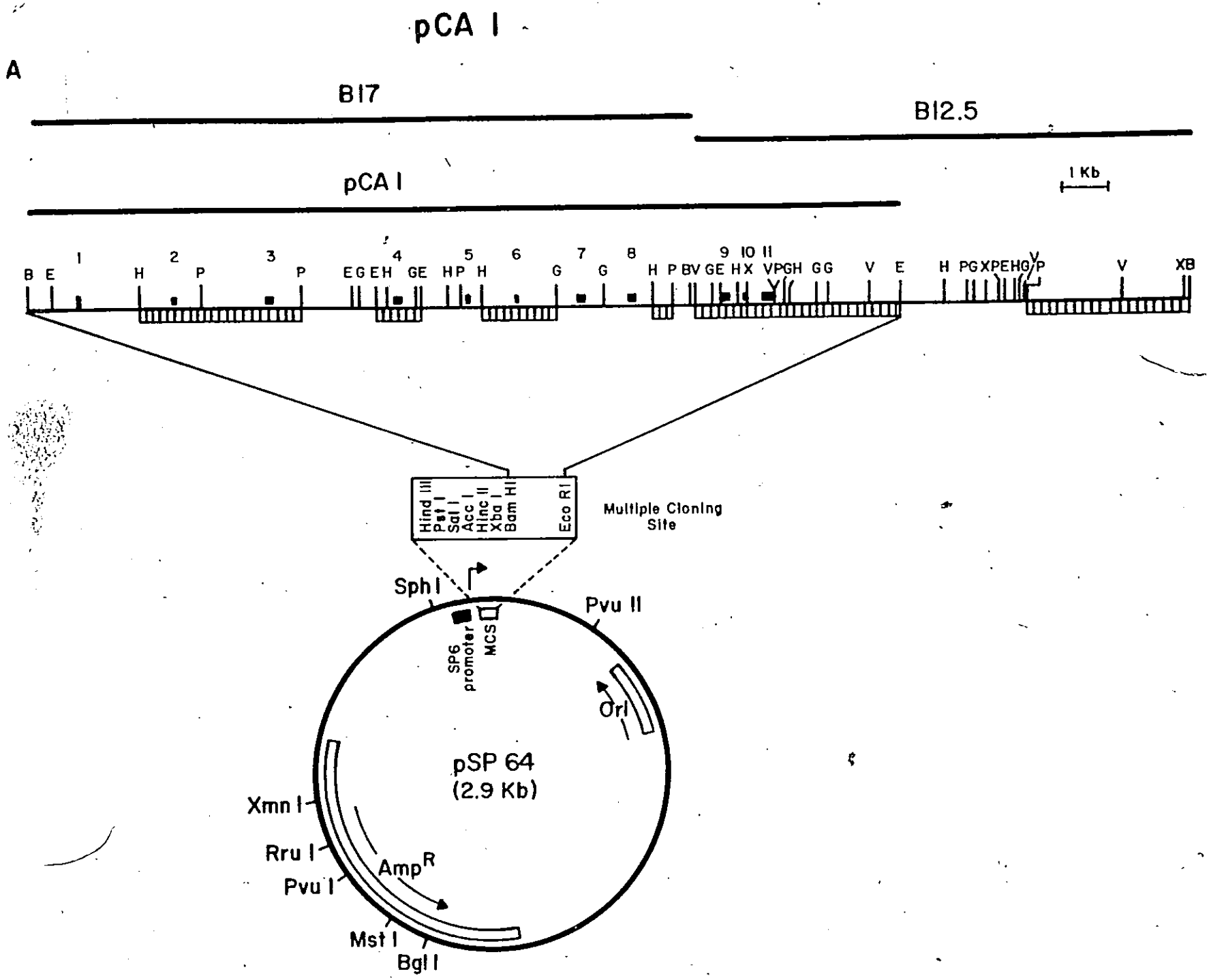
The recombinant pCA1pa-64. It contains exons 1 to 8 of the mouse pgk-1^b allele (Adra et al., 1987). This construct was used to subclone the 5'-flanking region and first exon of the mouse pgk-1 gene for nucleotide sequence analysis and expression studies. It was also used in the reconstruction of the pgk-1^b gene (see pCA1). The recombinant lambda genome carrying the B17 fragment (Adra et al., 1988) was digested to completion with BamHI then the insert was subcloned into the BamHI site of pSP64. Filled boxes represent exons; the striped boxes represent restriction fragments which contained repeated sequences. The meaning of the letters is as described in pCA2P-64 legend.

pCA I pa-64


B17
B E
1
H
2
P
3
P
EG EH
4
GE
HP
5
H
6
G
7
G
8
H P B
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Bam HI
Ava I
Sma I
Sac I
Eco RI
Multiple Cloning Site
Sph I
Pvu II
SP6 promoter
MCS
Ori
pSP64
(2.9 Kb)
Xmn I
Rru I
Pvu I
Amp^R
Mst I
Bgl I

APPENDIX 2 Cont'd

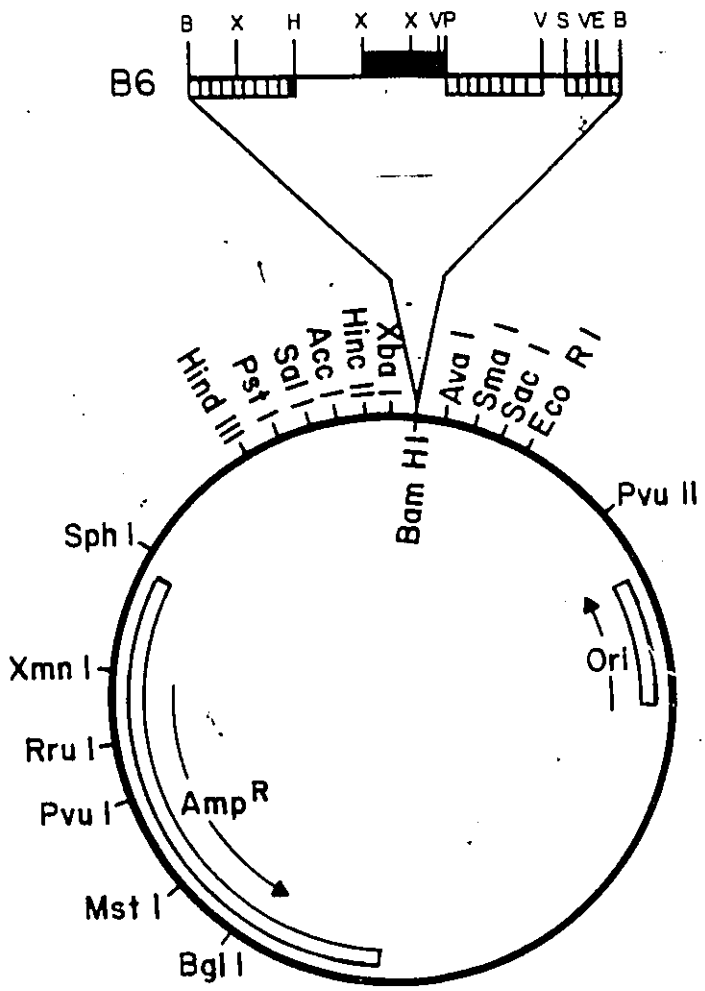
The recombinant pCA1. It contained the entire pgk-1^b gene (Adra et al., 1987) as well as 0.9 Kb of 5'-flanking sequence and 2.8 Kb of 3'-flanking sequence. It was constructed by digesting pCA1pa-64 (Appendix 2) with BamHI then the B17 insert was subcloned into the BamHI site of pCAB12.5BamHI-EcoRI4.5-64 (Appendix 2). Filled boxes represent exons; the striped boxes represent restriction fragments which contained repeated sequences. The meaning of the letters is as described in pCA2P-64.


pCA I
A
B17
B12.5
pCA I
1 Kb
B E 1 H 2 P 3 P E G E H 4 GE H P 5 H 6 G 7 G 8 H P B V GE H X V P G H 9 10 11 GG V E H P G X P E H G P V V XB
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Bam HI
Eco RI
Multiple Cloning Site
Sph I
Pvu II
SP6 promoter
MCS
Ori
pSP 64
(2.9 Kb)
Xmn I
Rru I
Pvu I
Amp R
Mst I
Bgl I

APPENDIX 2 Cont'd

The recombinant pCAB6-64. It contains the B6 fragment (Adra et al., 1988). The recombinant lambda genome carrying the B6 fragment was digested with BamHI then the insert was subcloned into the BamHI site of pSP64. The meaning of the symbols is as described in pCA2P-64 legend.

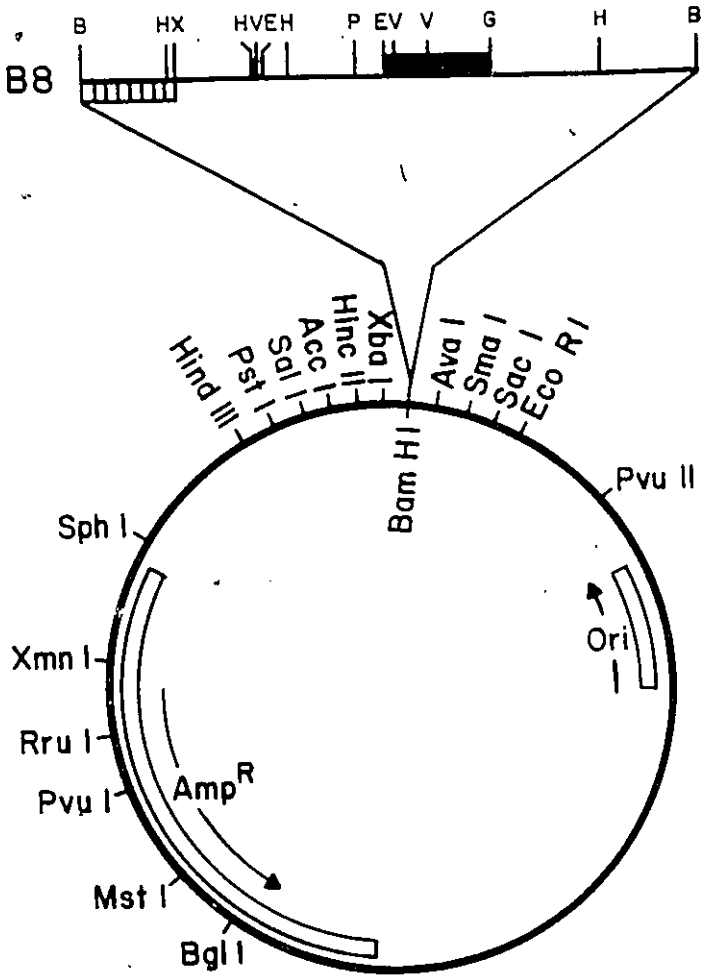
pCA B6-SP64


B6
B
X
H
X
X
VP
V
S
VE
B
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Ava I
Sma I
Sac I
Eco R I
Bam HI
Pvu II
Sph I
Xmn I
Rru I
Pvu I
Mst I
Bgl I
Amp R
Ori

APPENDIX 2 Cont'd

The recombinant pCAB8-64. It contains the B8 fragment (Adra et al., 1988). The recombinant lambda genome carrying the B8 fragment was cleaved with BamHI then the insert was ligated to the BamHI-cleaved pSP64. The meaning of the symbols is as described in the pCA2P-64 legend.

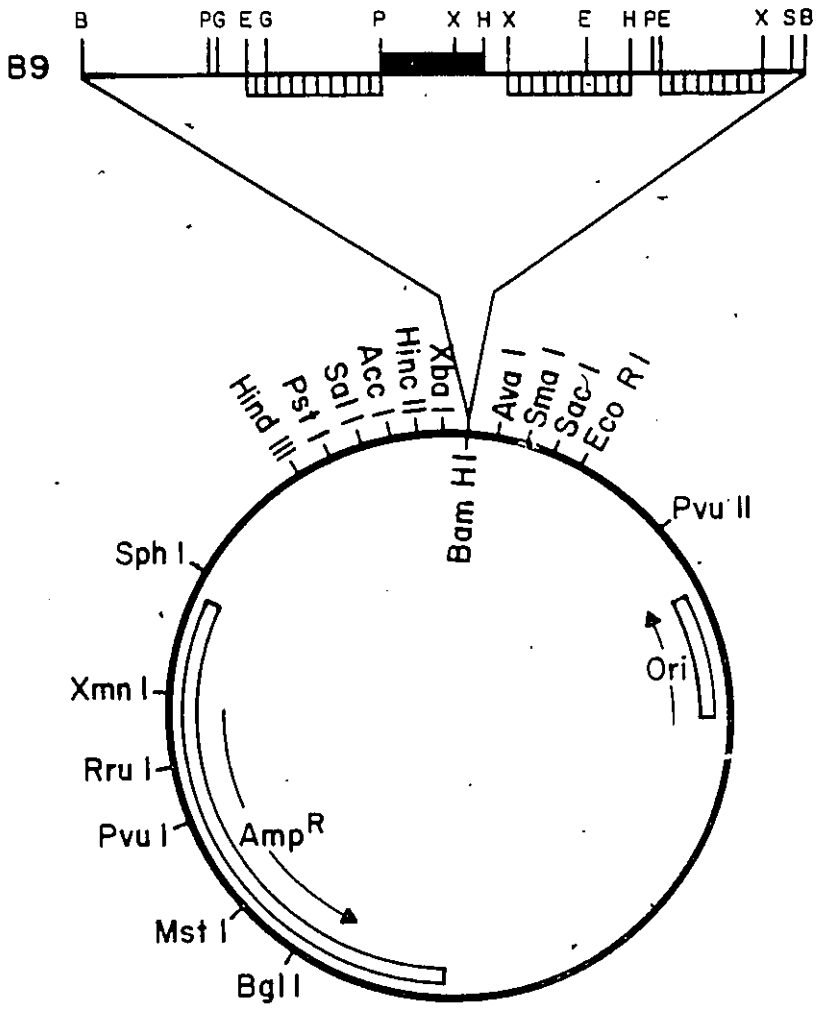
pCA B8-SP64


B8
B
HX
HVEH
P
EV
V
G
H
B
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Bam HI
Ava I
Sma I
Sac I
Eco RI
Pvu II
Sph I
Xmn I
Rru I
Pvu I
Mst I
Bgl I
Amp^R
Ori

APPENDIX 2 Cont'd

The recombinant pCAB9-64. It contains the B10 fragment (Adra et al., 1988). The recombinant lambda genome carrying the B10 fragment was digested with BamHI then the insert was subcloned into the BamHI site of pSP64. The meaning of the symbols is as described in pCA2P-64 legend.

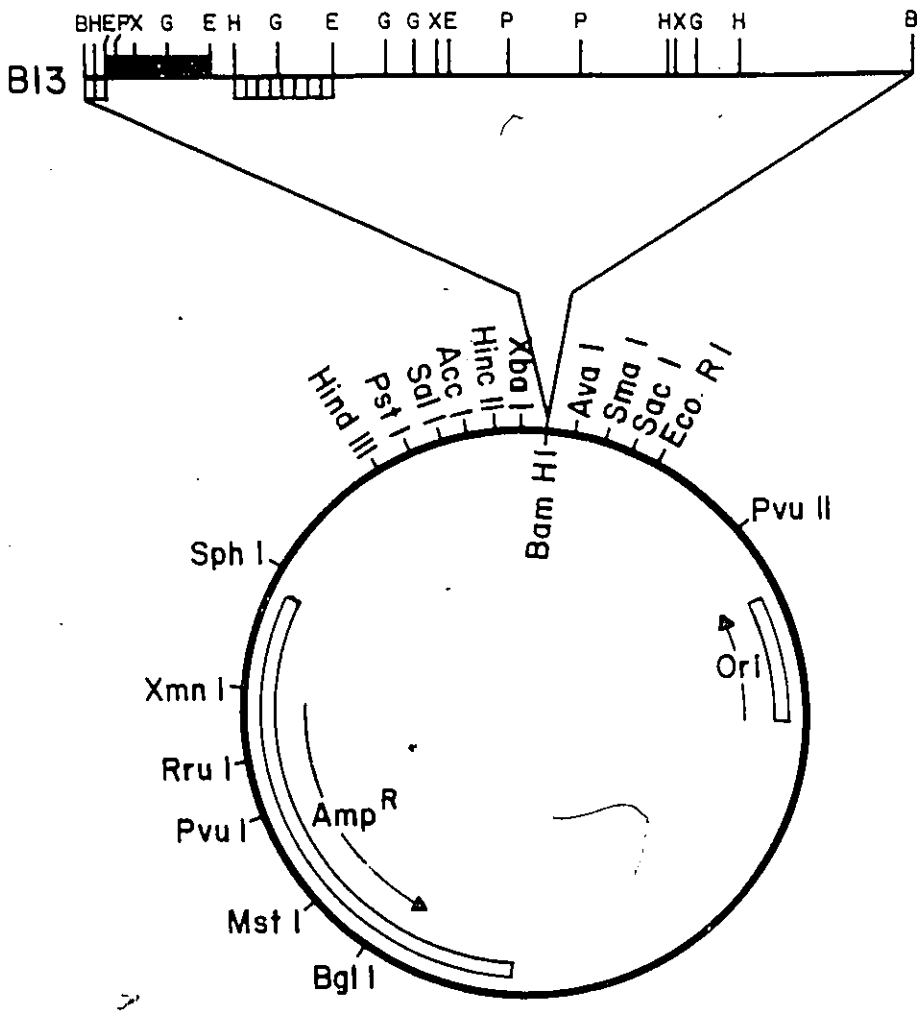
pCA B9-SP64


B9
B
PG
E
G
P
X
H
X
E
H
P
E
X
S
B
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Bam HI
Ava I
Sma I
Sac I
Eco RI
Pvu II
Sph I
Xmn I
Rru I
Pvu I
Mst I
Bgl I
Amp^R
Ori

APPENDIX 2 Cont'd

The recombinant pCAB13-64. It contains the B13 fragment (Adra et al., 1988). The recombinant lambda genome carrying the B13 fragment was digested with BamHI then the insert was subcloned into the BamHI site of pSP64. The meaning of the symbols is as described in pCA2P-64 legend.

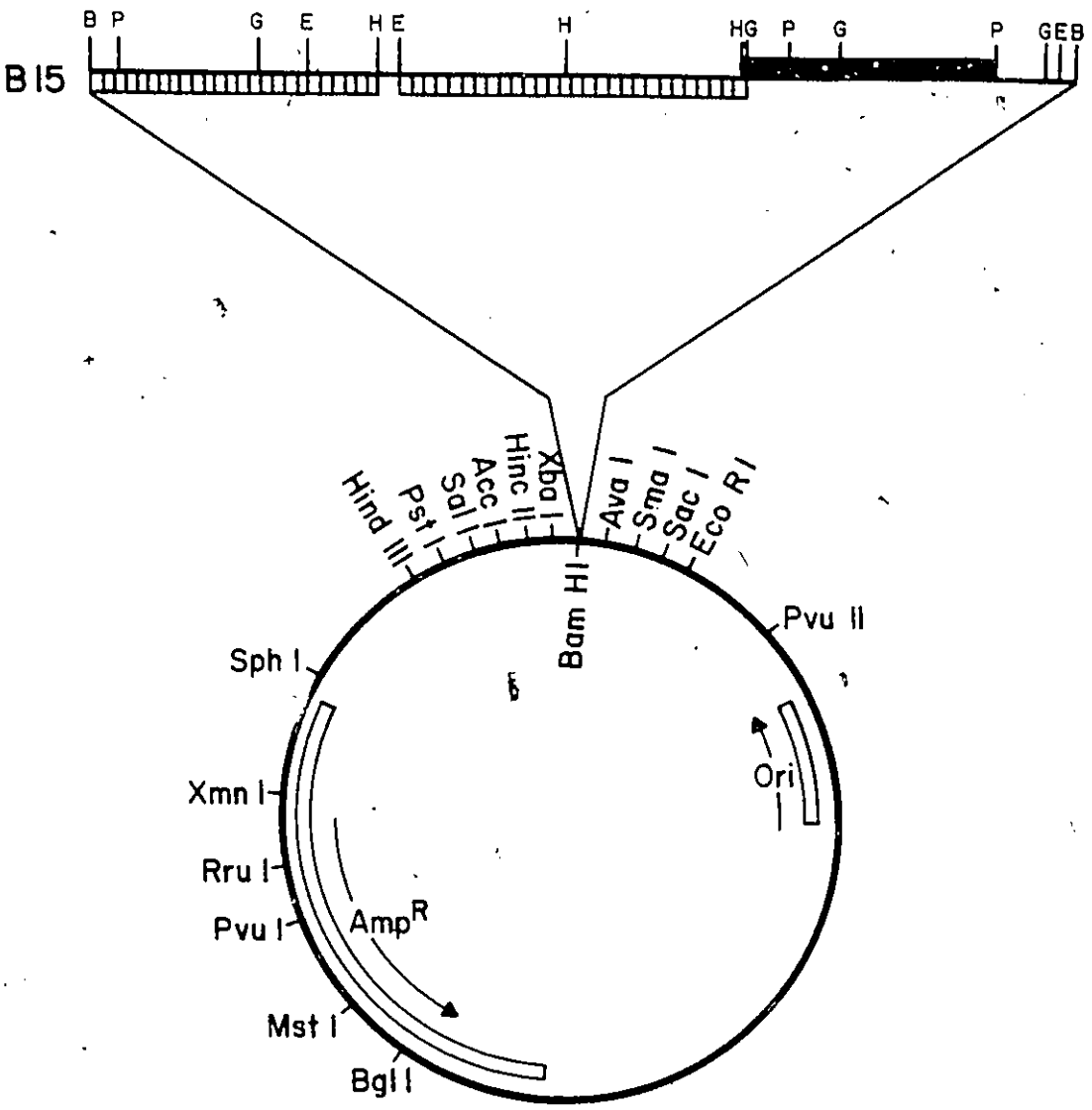
pCA B13-SP64


B13
B H E P X G E H G E G G X E P P H X G H B
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Bam HI
Ava I
Sma I
Sac I
Eco RI
Pvu II
Sph I
Xmn I
Rru I
Pvu I
Mst I
Bgl I
Amp R
Ori

APPENDIX 2 Cont'd

The recombinant pCAB15-64. It contains the B15 fragment (Adra et al., 1988). The recombinant lambda genome carrying the B15 fragment was digested with BamHI then the insert was subcloned into the BamHI site of pSP64. The meaning of the symbols is as described in pCA2P-64 legend.

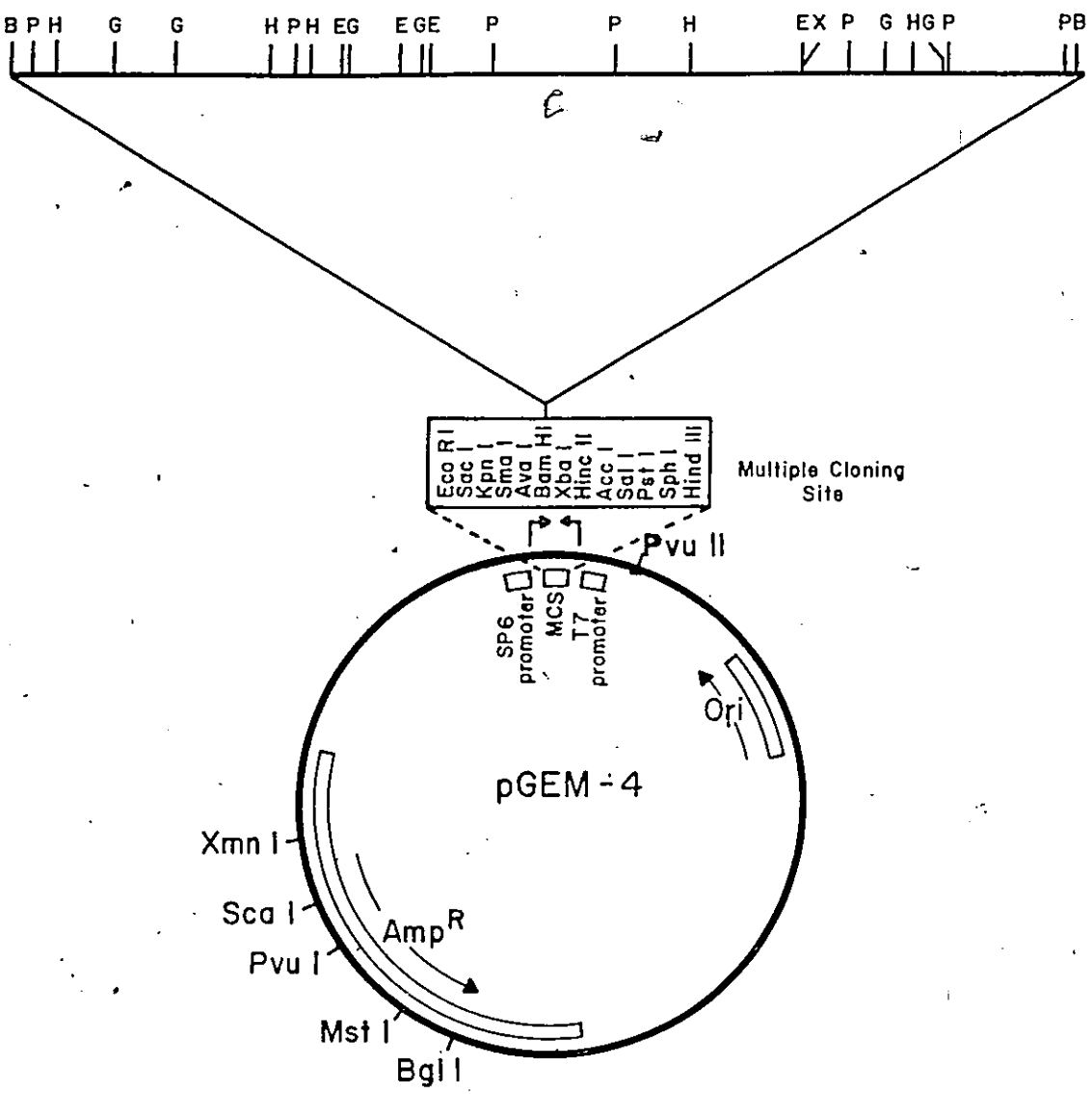
pCA B15-SP64


B15
B P G E H E H HG P G P G E B
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Bam HI
Ava I
Sma I
Sac I
Eco RI
Pvu II
Ori
Sph I
Xmn I
Rru I
Pvu I
Mst I
Bgl I
AmpR

APPENDIX 2 Cont'd

The recombinant pCAB21-GEM-4. It contains the first 8 exons of pgk-1^{a} (see Chapter III). The recombinant lambda genome carrying the fragment B21 (Adra et al., 1988) was digested with BamHI then the insert was subcloned into the BamHI site of GEM-4. The meaning of the letters is as described in pCA2P-64 legend.

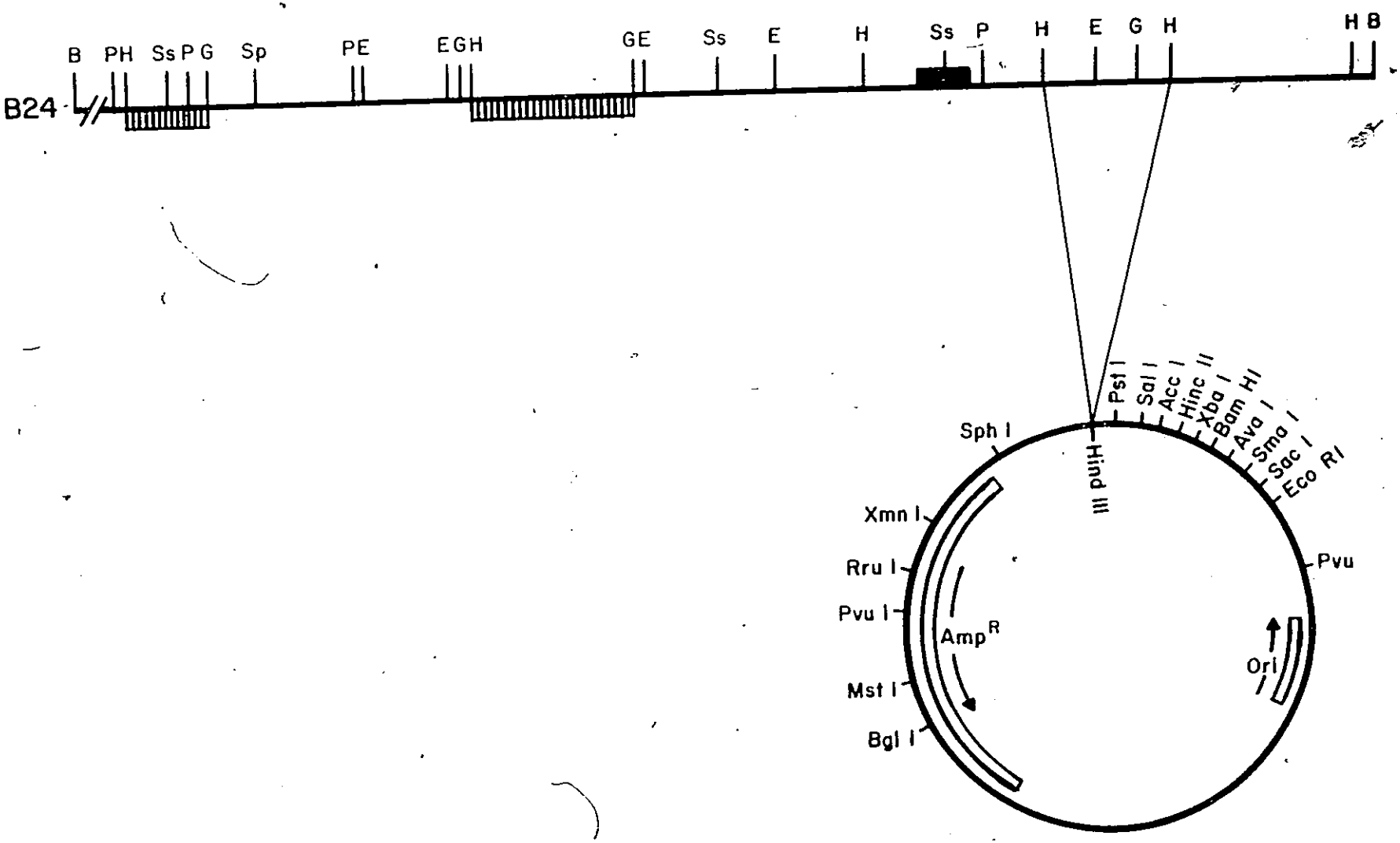
pCA B21-GEM-4


B P H G G H P H EG E GE P P H EX P G HG P PB
Eco RI
Sac I
Kpn I
Sma I
Ava I
Bam HI
Xba I
Hinc II
Acc I
Sal I
Pst I
Sph I
Hind III
Multiple Cloning Site
Pvu II
SP6 promoter
MCS
T7 promoter
Ori
pGEM-4
Xmn I
Sca I
Pvu I
Mst I
Bgl I
Amp^R

APPENDIX 2 Cont'd

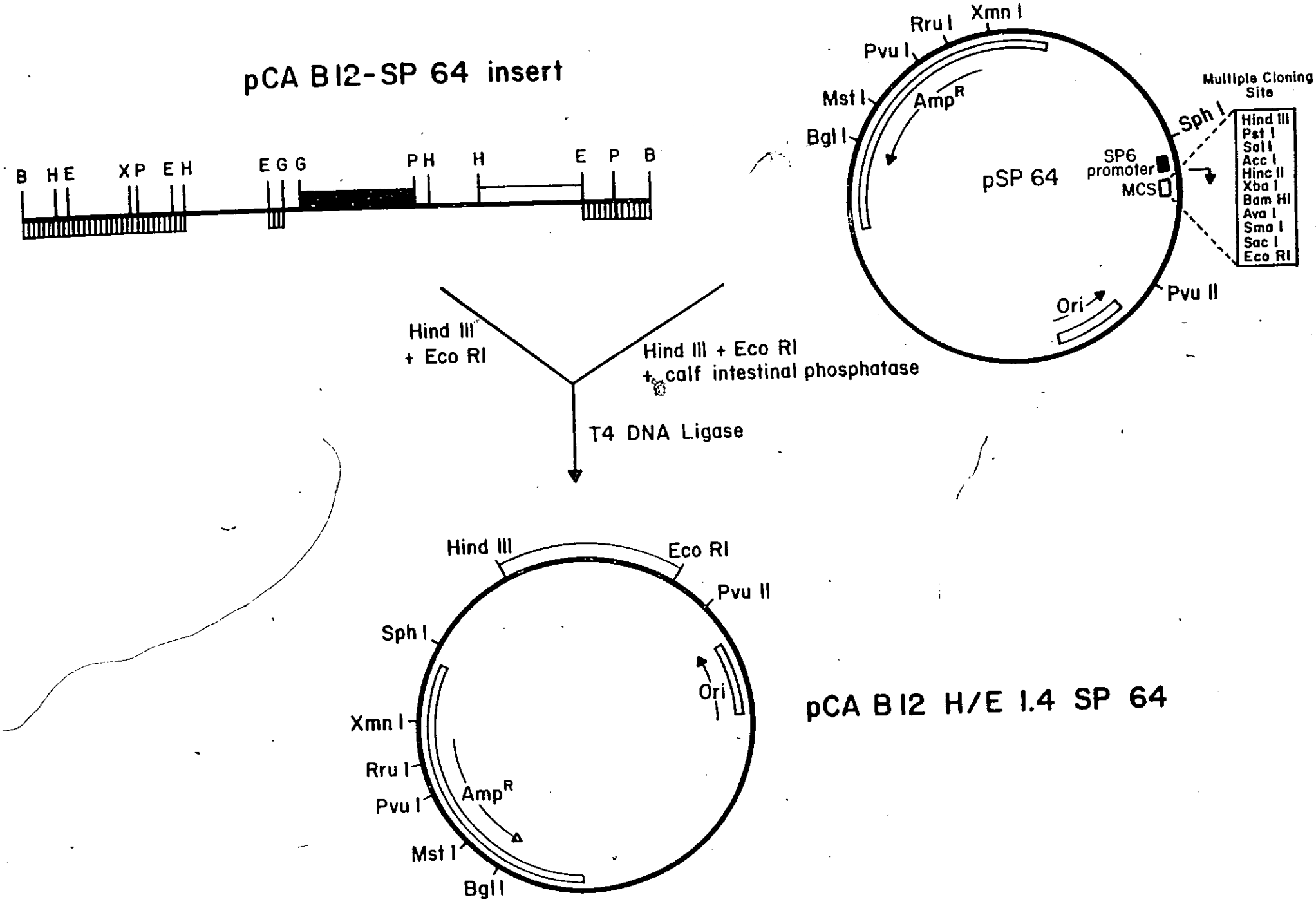
The recombinant pB24H1.7-SP64 contains a unique fragment derived from B24. It was used to determine the chromosomal location of B24 (Adra et al., 1988). A 1.7-Kb HindIII fragment to the right of the pgk-related sequence was subcloned in the HindIII site of pSP64. The meaning of the various symbols is as described in pCA2P-64 legend.

pB24 H 1.7-SP64



APPENDIX 2 Cont'd

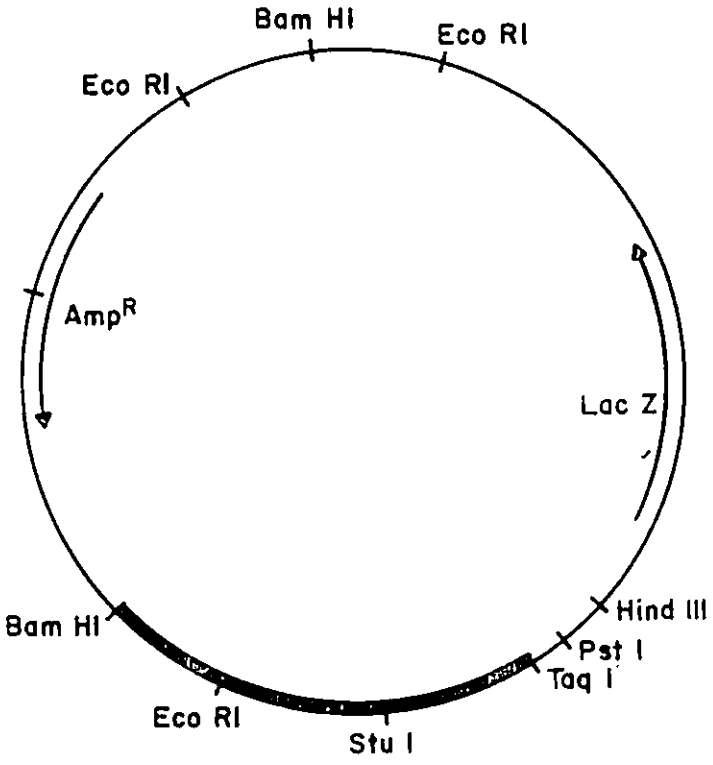
The construct pCAB12H/E1.4-SP64 contains a unique fragment derived from B12 (Adra et al., 1988). It was used to map B12 in somatic cell hybrids. A 1.4-Kb HindIII/EcoRI fragment to the left of the pgk-2 gene was subcloned in the HindIII/EcoRI-cleaved pSP64. The meaning of the various symbols is as described in pCA2P-64 legend.


pCA B12-SP 64 insert
B H E X P E H E G G P H H E P B
Rru I
Xmn I
Pvu I
Mst I
Amp^R
Bgl I
Sph I
Multiple Cloning Site
Hind III
Pst I
Sal I
Acc I
Hinc II
Xba I
Bam HI
Ava I
Sma I
Sac I
Eco RI
SP6 promoter
MCS
pSP 64
Ori
Pvu II
Hind III + Eco RI
Hind III + Eco RI + calf intestinal phosphatase
T4 DNA Ligase
Hind III
Eco RI
Pvu II
Sph I
Ori
Xmn I
Rru I
Amp^R
Pvu I
Mst I
Bgl I
pCA B12 H/E 1.4 SP 64

APPENDIX 2 Cont'd

The construct pCAB17prBamHI-TaqI919LacZ. It contains the BamHI/TaqI fragment derived from the 5'-end of the pgk-1^b allele (Adra et al., 1987). This construct was prepared by digesting pCAB17prBamHI/TaqI 919-GEM-4 with BamHI/HindIII (Appendix 2), the BamHI/HindIII fragment was then subcloned into pCH126A2 which was digested to completion with HindIII and partially with BamHI.

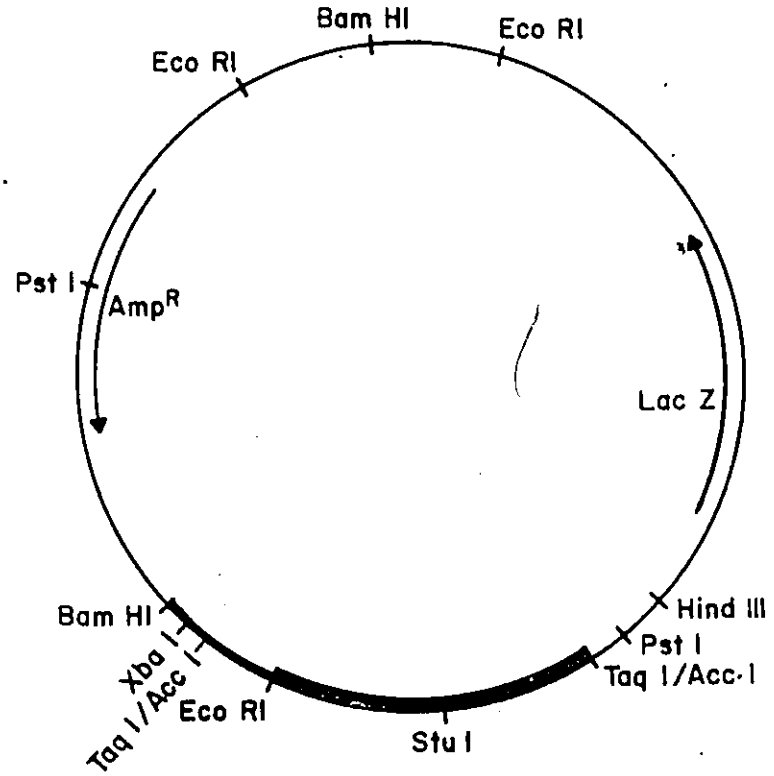
pCA B17 pr Bam HI- Taq I 919 Lac Z



APPENDIX 2 Cont'd

The recombinant pCAB17prEcoRI-TaqI500LacZ. It contains the EcoRI/TaqI fragment (indicated by a filled box) derived from the 5' end of the pgk-1^b allele (Adra et al., 1987). This construct was prepared by digesting pCAB17pr1100-GEM-4 with BamHI/HindIII. The BamHI/HindIII fragment was then subcloned into HindIII (partial)/ BamHI-cleaved pCH126A2.

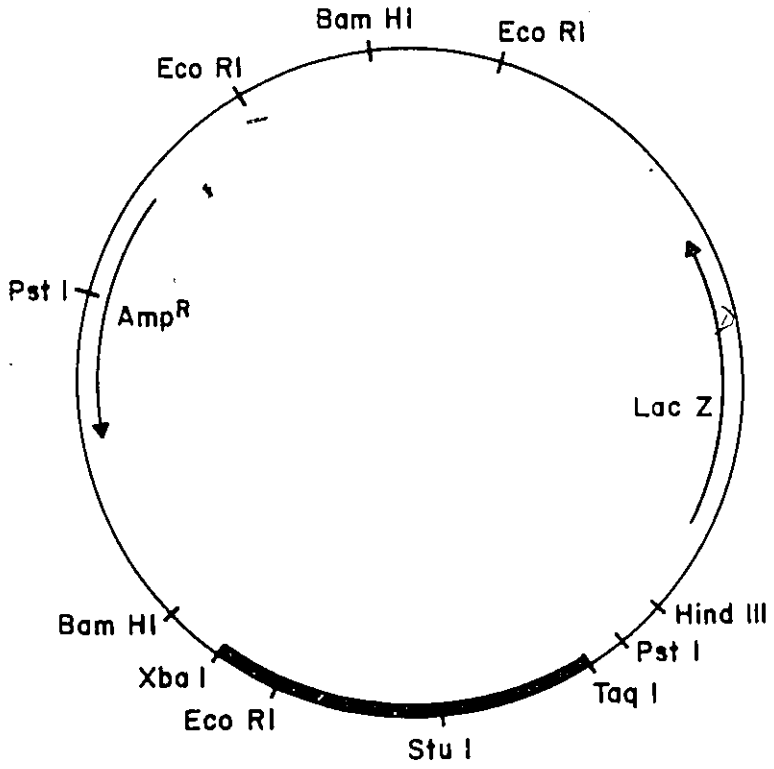
pCA 817 pr Eco RI - Taq I 500 Lac Z



APPENDIX 2 Cont'd

The recombinant pCAB17prXbaI-TaqI520LacZ. It contains a 520-bpXbaI-TaqI fragment (indicated by a filled box) derived from the 5' end of the pgk-1^b allele (Adra et al., 1987). This construct was prepared by digesting pCAB17prXbaI/TaqI520-GEM-4 (Appendix 2) with BamHI + HindIII. The BamHI/HindIII fragment which carries the XbaI-TaqI fragment was then subcloned into HindIII (partial)/BamHI-cleaved pCH126A2.

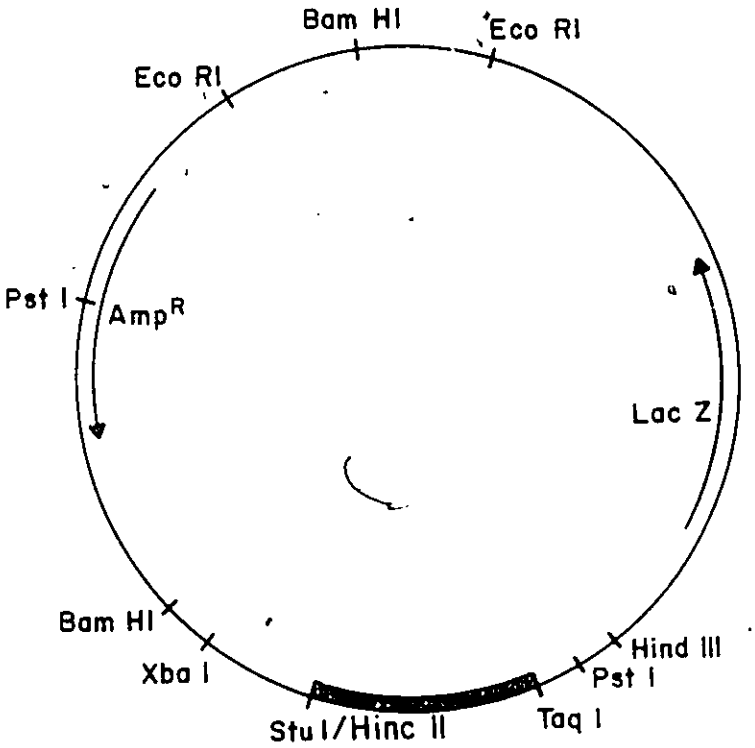
pCA BI7 pr Xba I-Taq I 520 Lac Z



APPENDIX 2 Cont'd

The recombinant pCAB17prStuI/TaqI200LacZ. It contains a 200-bp StuI/TaqI fragment (indicated by a filled box) derived from the 5'-flanking region of the pgk-1^b allele (Adra et al., 1987). This construct was constructed by digesting pCAB17prStuI/TaqI200-GEM-4 (Appendix 2) with BamHI + HindIII. The BamHI/HindIII fragment which carries the StuI/TaqI insert was then subcloned into HindIII (partial)/BamHI-cleaved pCH126A2.

pCA BI7 pr Stu I-Taq I 200 Lac Z

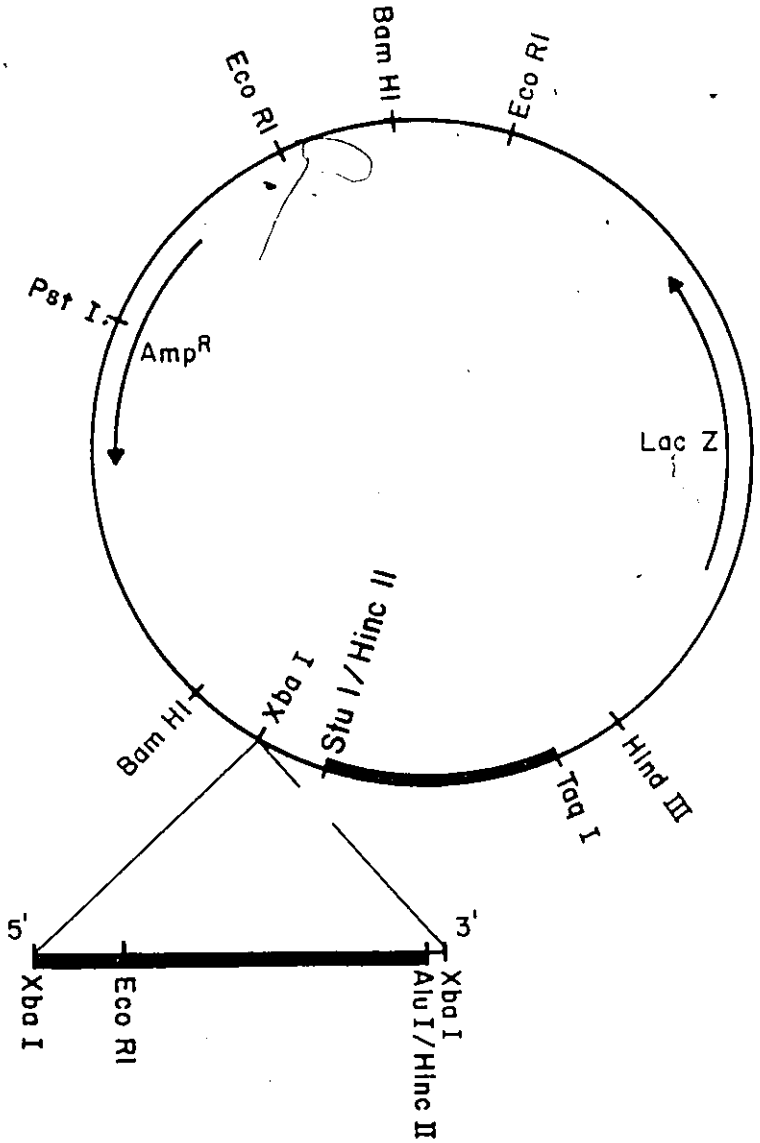


APPENDIX 2

The recombinant pCA1014. It was constructed by ligating the XbaI fragment derived from pCA1010 (Appendix 2) directly into pCAB17prStuI-TaqI200LacZ (Appendix 2) digested with XbaI and treated with calf intestinal phosphatase. The orientation of the DNA inserts was determined by restriction analysis.

pCA 1014

pCA B17 Xba I-Alu I B17 Stu I-Taq I Lac Z 5'→3'

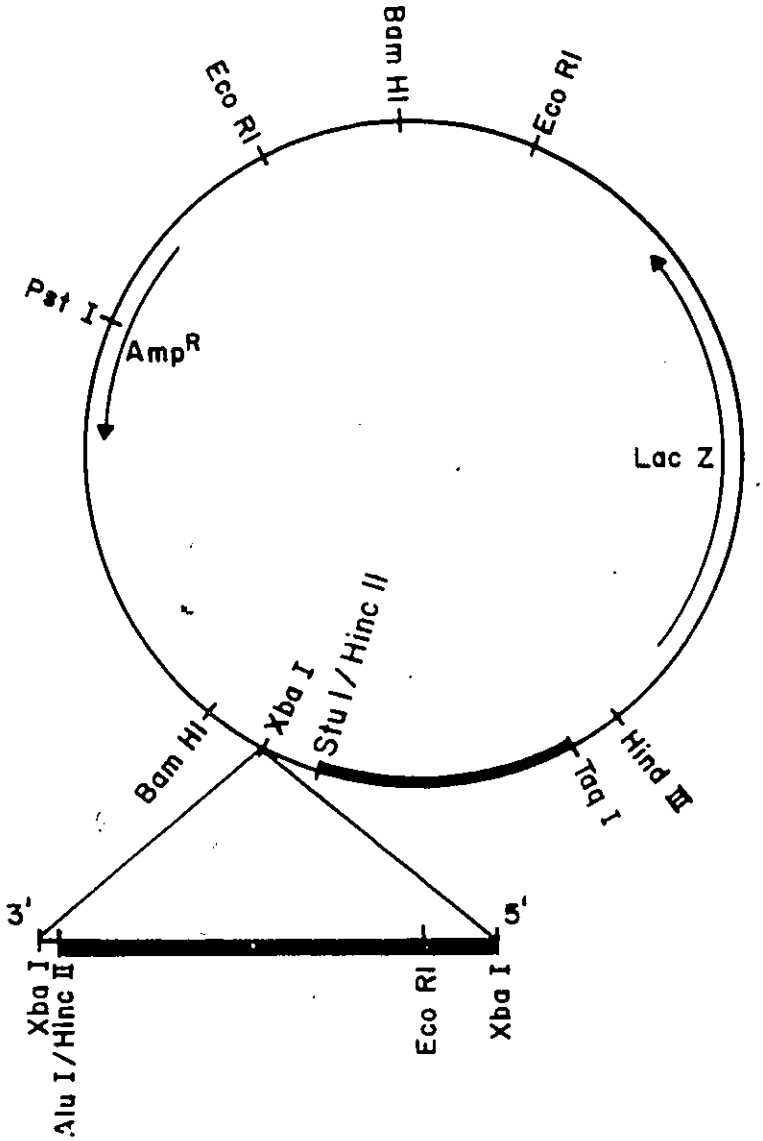


APPENDIX 2 Cont'd

The recombinant pCA1015. It contains the same insert as pCA1014 except that it is in the reverse orientation (see pCA1014 legend).

pCA 1015

pCA B17 Xba I-Alu I B17 Stu I-Taq I Lac Z 3'⟶5'

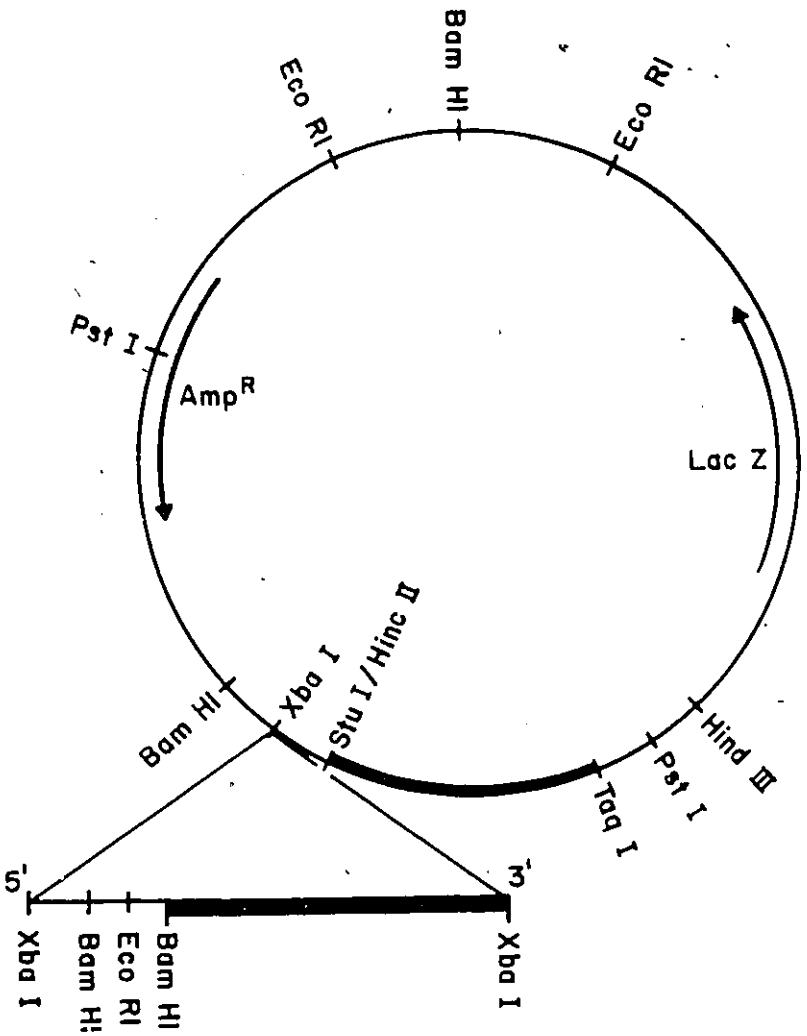


APPENDIX 2 Cont'd

The recombinant pCA1016. It was constructed by ligating the 4.5-Kb XbaI fragment derived from pCA1001 (Appendix 2) into pCAB17prStuI-TaqI200LacZ (Appendix 2) digested with XbaI and treated with calf intestinal phosphatase. The orientation of the DNA inserts was determined by restriction analysis.

pCA 1016

pCA B21 Bam HI-Xba I 4.5 B17 Stu I-Taq I Lac Z 5'⟶3'

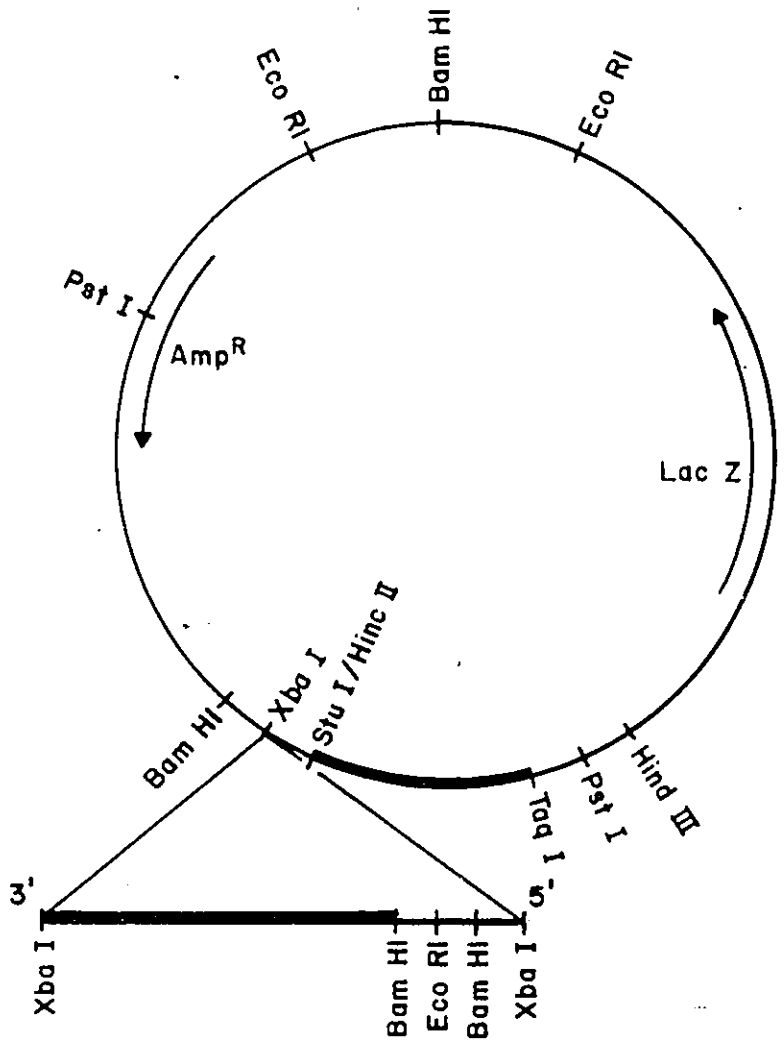


APPENDIX 2 Cont'd

The recombinant pCA1016'. It contains the same insert as pCA1016 except that it is in the reverse orientation (see pCA1016 legend).

pCA 1016'

pCA B21 Bam HI-Xba I 4.5 B17 Stu I-Taq I Lac Z 5'⟶3'

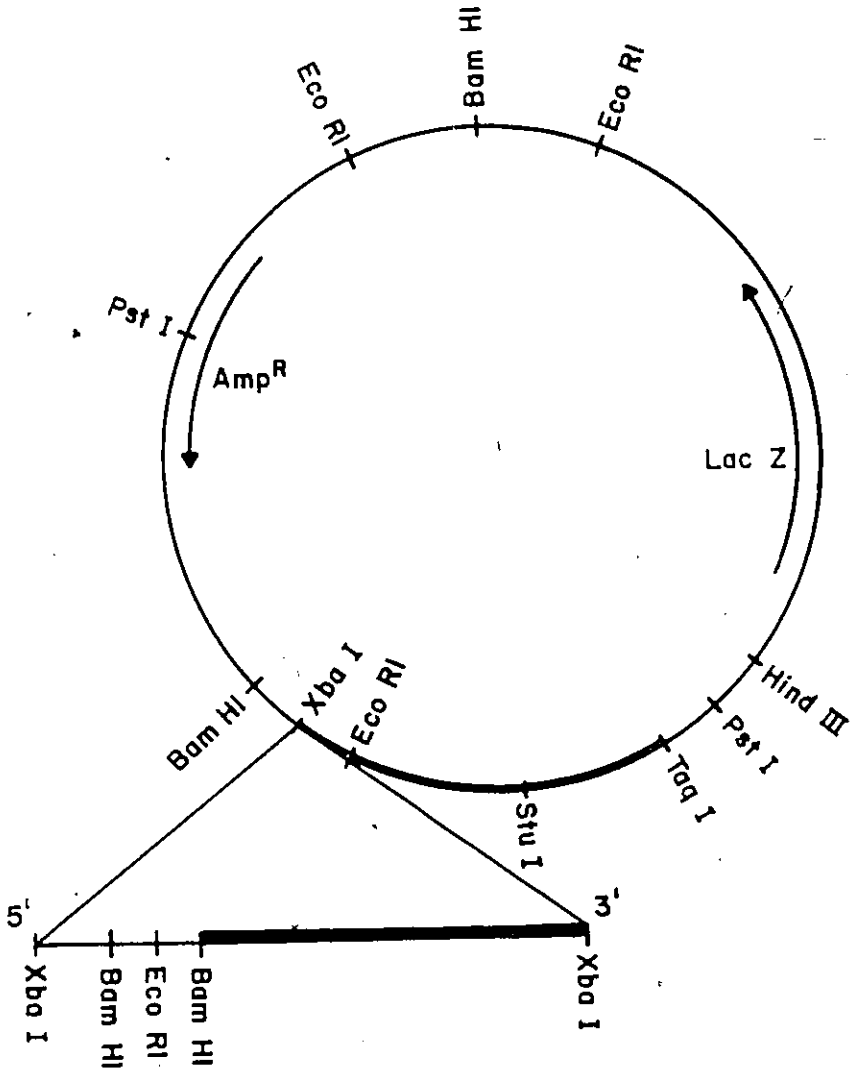


APPENDIX 2 Cont'd

The recombinant pCA1017. It was constructed by ligating the XbaI fragment derived from pCA1001 (Appendix 2) into pCAB17prXbaI-TaqI520 LacZ (Appendix 2) digested with XbaI and treated with calf intestinal phosphatase. The orientation of the insert was determined by restriction analysis.

pCA 1017

pCA B21 Bam HI-Xba I 4.5 B17 Xba I-Taq Lac Z, 5'⟶3'

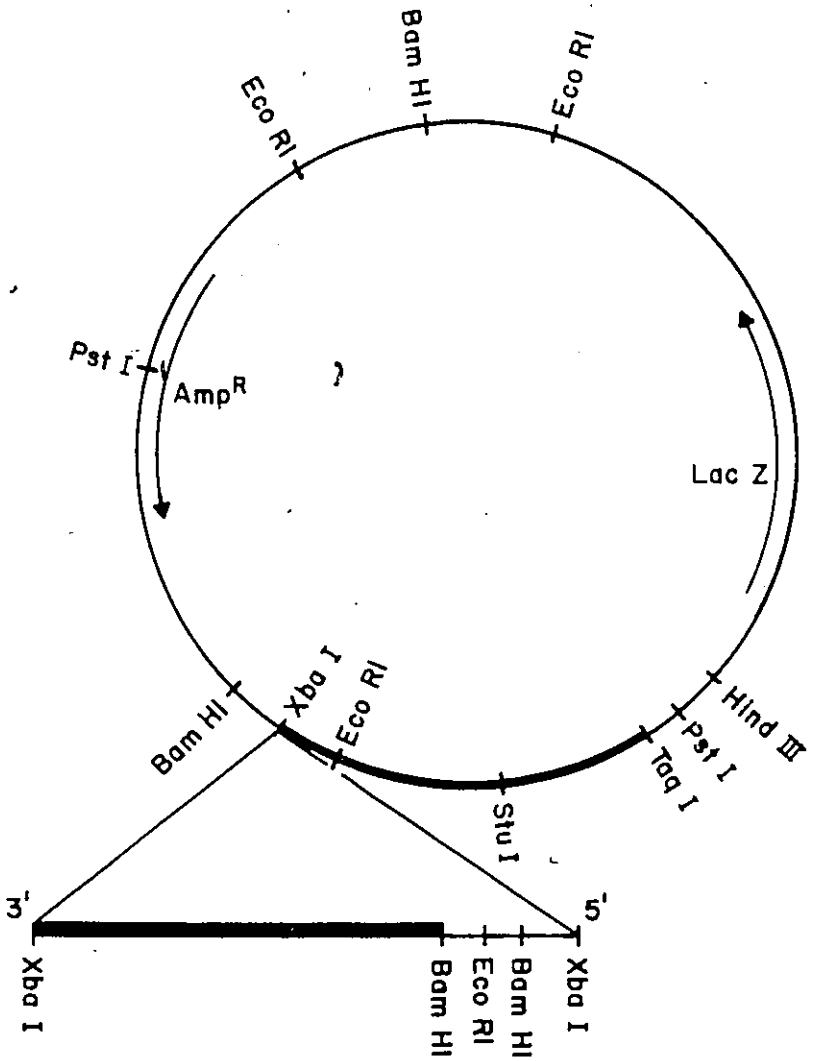


APPENDIX 2 Cont'd

The recombinant pCA1018. It contains the same insert as pCA1017 (Appendix 2) except that it is in the reverse orientation (see pCA1017 legend).

pCA 1018

pCA B21 Bam HI-Xba I 4.5 B17 Xba I-Taq I Lac Z 3'⟶5'

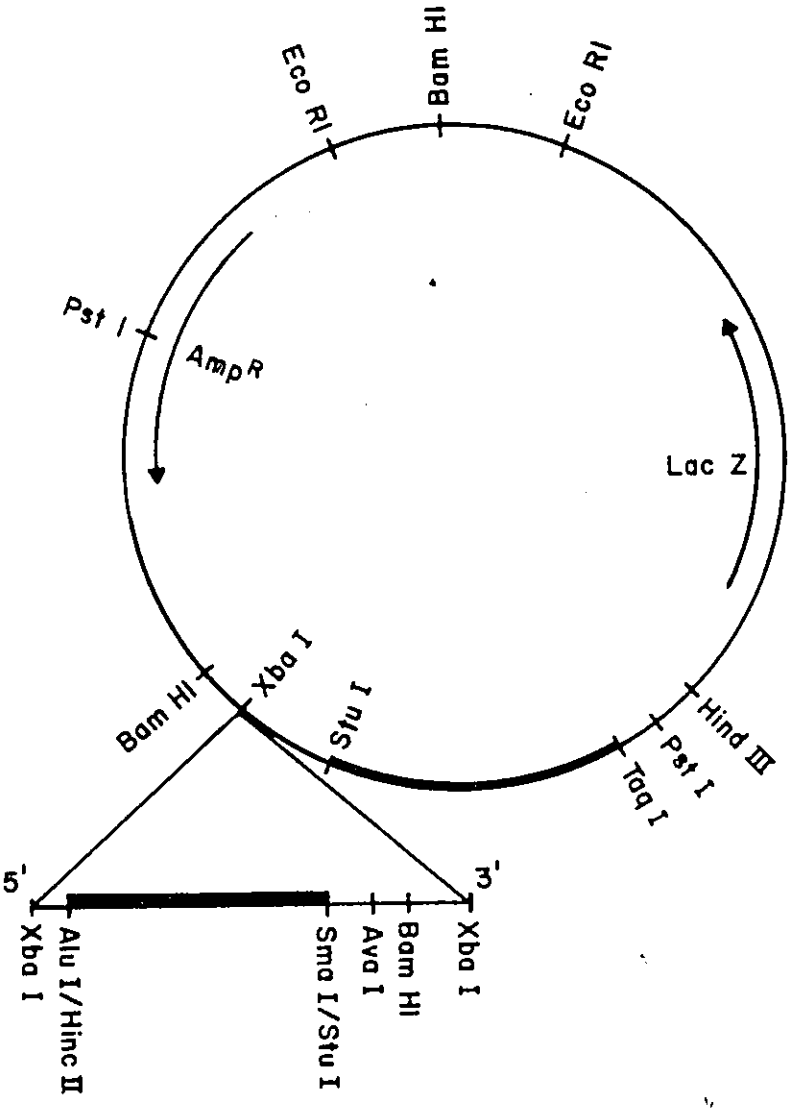


APPENDIX 2 Cont'd

The recombinant pCA1022. This construct was prepared by ligating the XbaI fragment derived from pCA1019 (Appendix 2) into pCAB17prStuI-TaqI200LacZ (Appendix 2) digested with XbaI and treated with calf intestinal phosphatase. The orientation of the insert was determined by restriction analysis.

pCA 1022

pCA B17 Alu I-Stu I 92 B17 Stu I-Taq I Lac Z 5'⟶3'

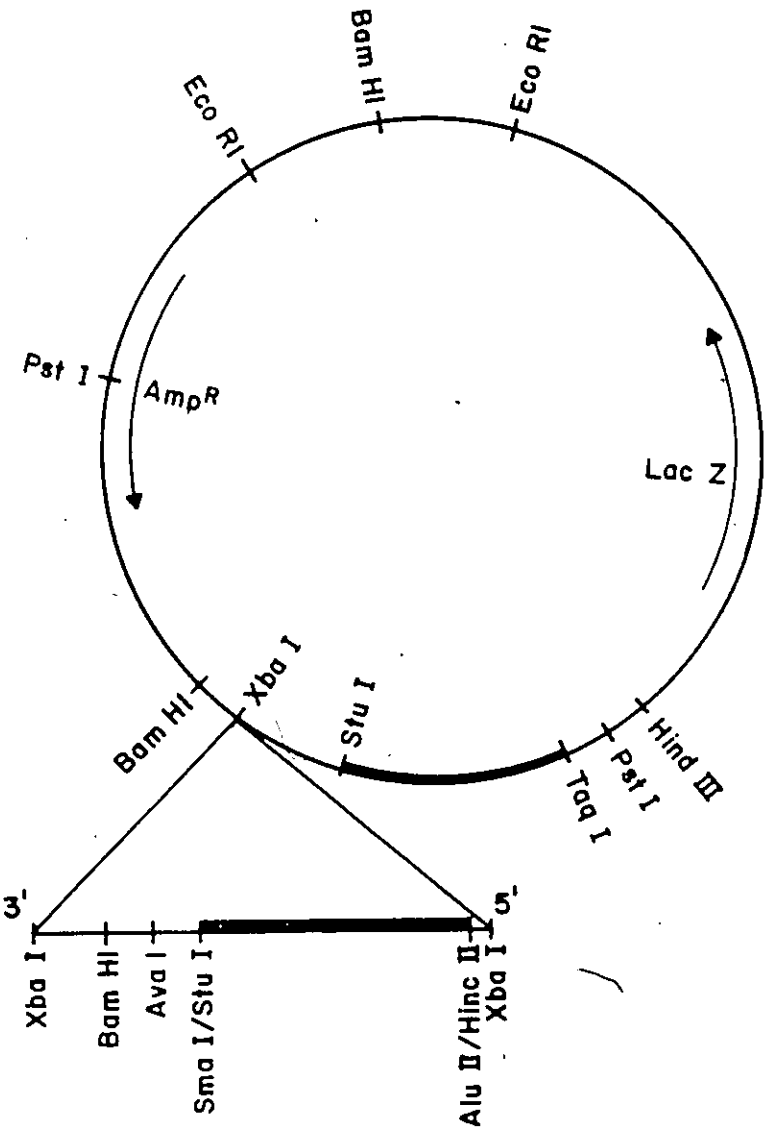


APPENDIX 2 Cont'd

The recombinant pCA1023. It contains the same insert as pCA1022 (Appendix 2) except that it is in the reverse orientation (see pCA1022 legend).

pCA 1023

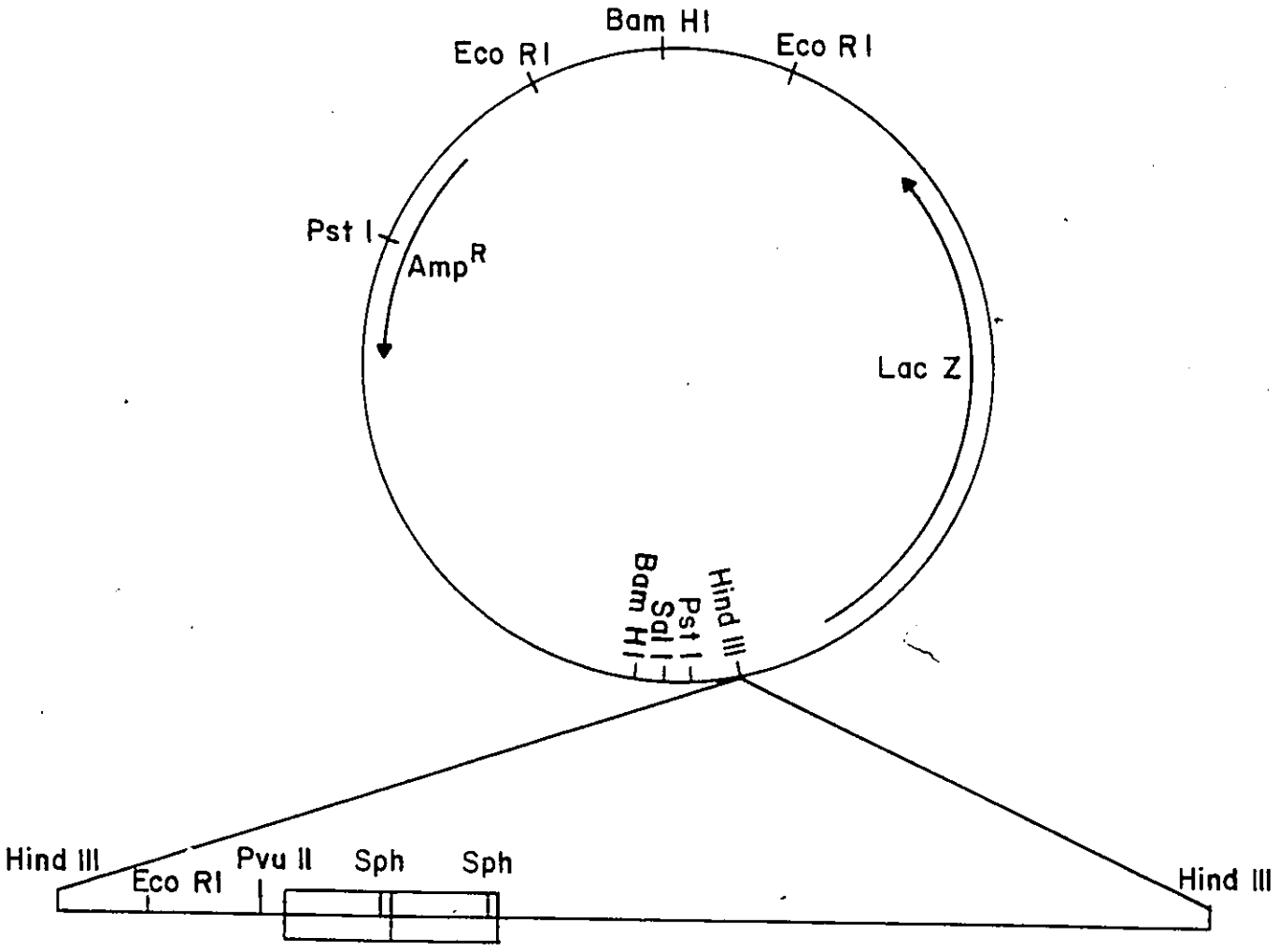
pCA BI7 Alu I-Stu I 92 BI7 Stu I-Taq I Lac Z 3'⟶5'



APPENDIX 2 Cont'd

The recombinant pCA1026. It carries the LacZ gene under the control of the SV40 promoter and its enhancers. This construct was prepared by ligating the 440-bp HindIII fragment derived from pSV1BSS81 into pCH126A2 digested with HindIII and treated with calf intestinal phosphatase.

pCA 1026

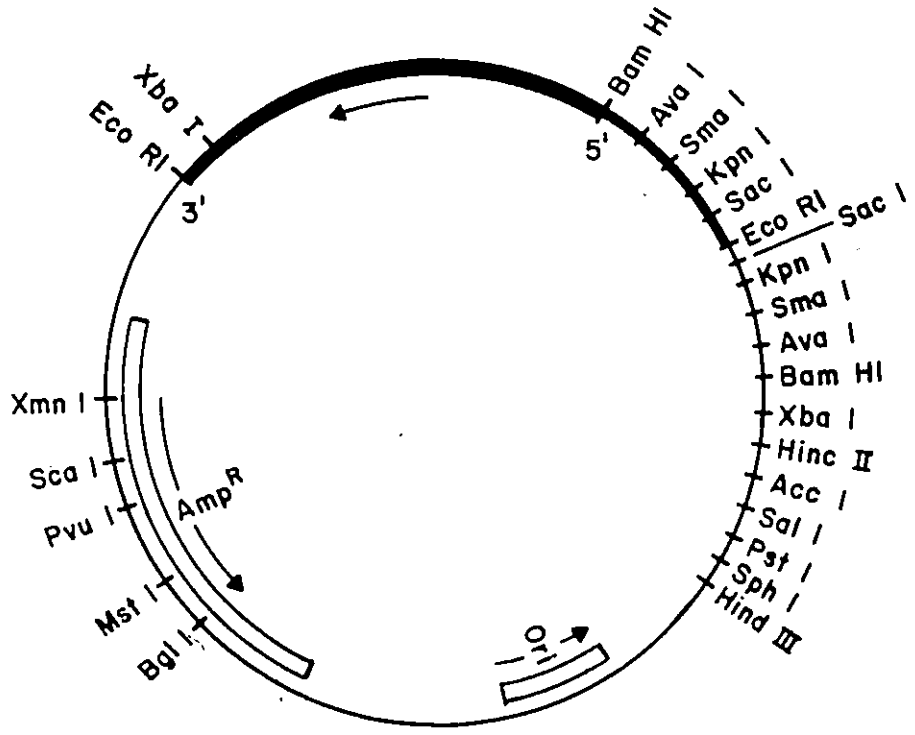

Bam HI
Eco RI
Eco RI
Pst I
Amp^R
Lac Z
Bam HI
Sal I
Pst I
Hind III
Hind III
Eco RI
Pvu II
Sph
Sph
Hind III

APPENDIX 2 Cont'd

The recombinant pCA1001 or pCA B21 E 4.5 pr; G4. It contains the 5'-flanking sequence of the pgk-1^a gene (Chapter III). It was prepared by digesting the recombinant pCAB21-GEM-4 (Appendix 2) with EcoRI then the 4.5-Kb EcoRI fragment carrying the BamHI-EcoRI fragment (indicated by a filled box) was subcloned into the EcoRI site of Gemini 4.

pCA 1001

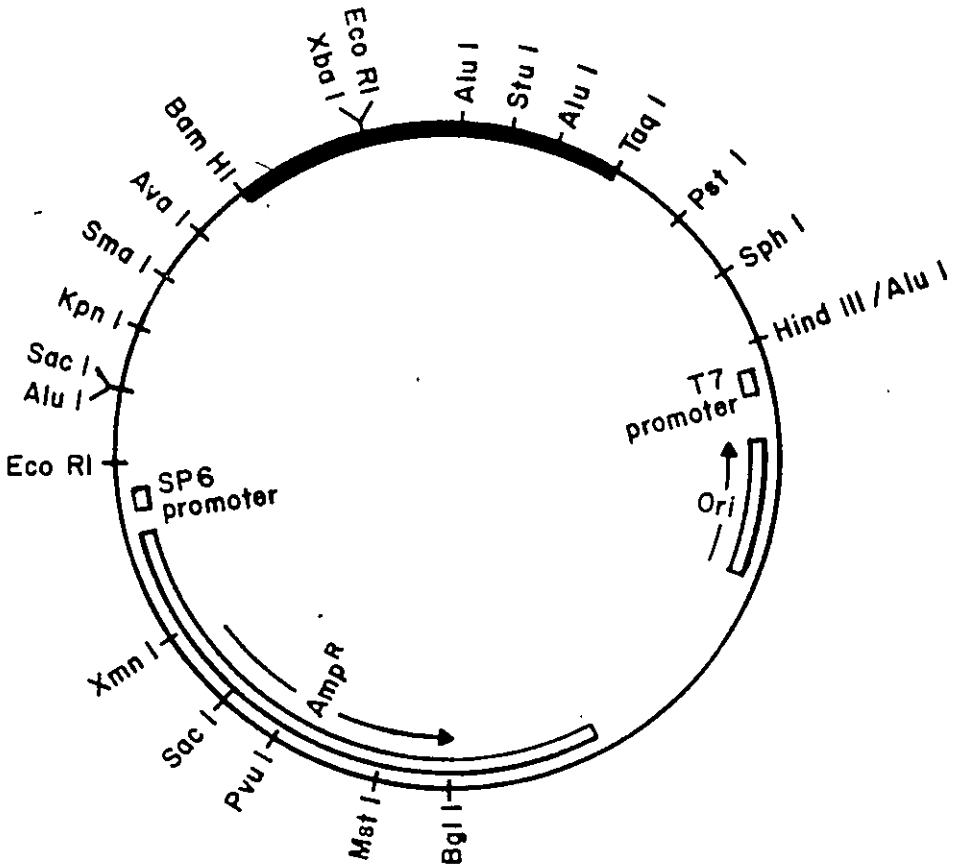
pCA B21 E 4.5 pr ; G4



APPENDIX 2 Cont'd

The recombinant pCAB17prBamHI/TaqI919-GEM-4. It contains a 919-bp BamHI/TaqI fragment derived from the 5'-flanking sequence of pgk-1^b gene (Adra et al., 1987). This construct was prepared by subcloning the BamHI/TaqI fragment (indicated by a filled box) into a BamHI/AccI-cleaved Gemini 4.

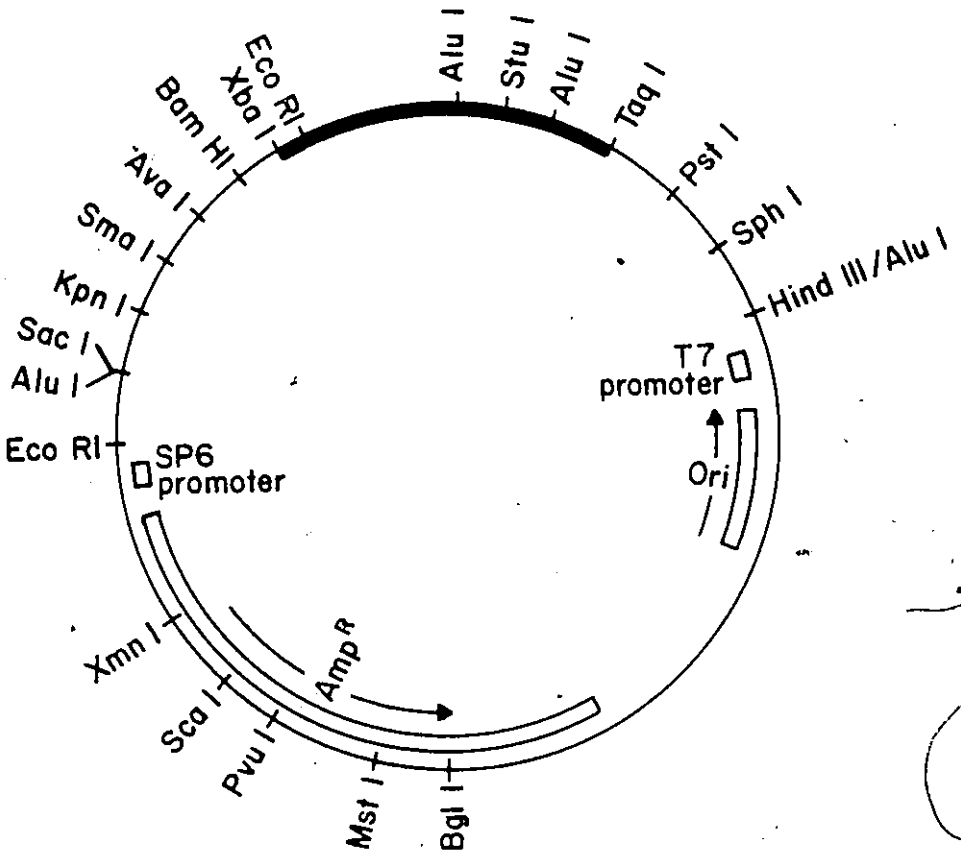
pCA B17 pr Bam HI / Taq I 919-GEM™-4



APPENDIX 2 Cont'd

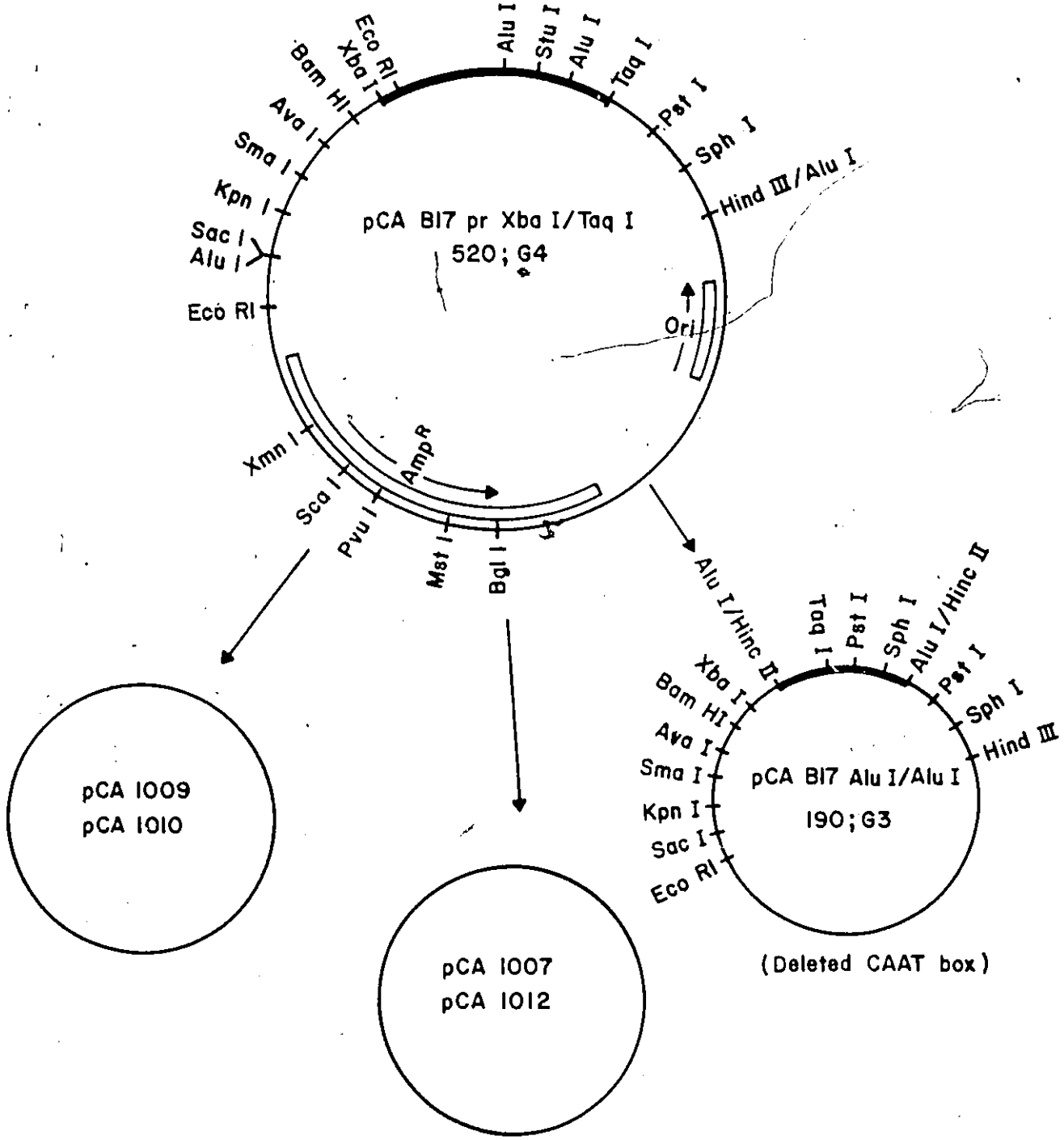
The recombinant pCAB17prXbaI/TaqI520-GEM-4. This construct contains a 520-bp XbaI/TaqI fragment derived from the 5'-flanking sequence of the mouse pgk-1^b allele (Adra et al., 1987). It was constructed by subcloning this XbaI/TaqI fragment (indicated by a filled box) into a XbaI/AccI-cleaved Gemini 4.

pCA B17 pr Xba I / Taq I 520-GEM™-4



APPENDIX 2 Cont'd

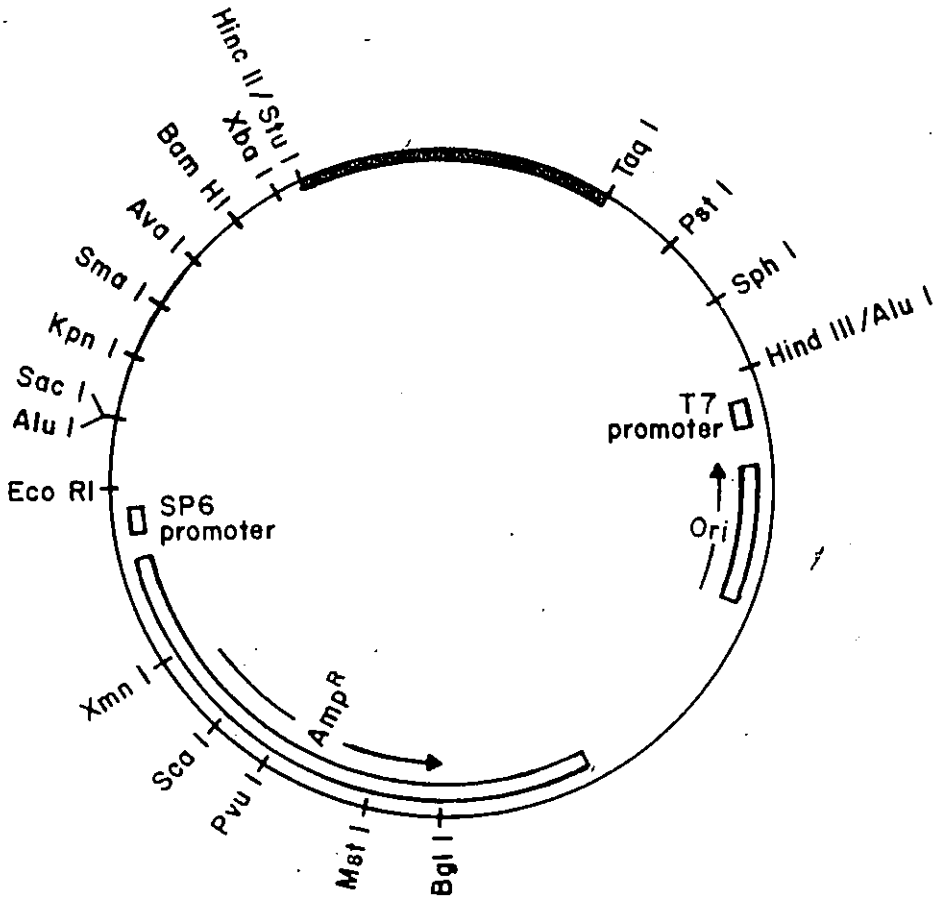
These graphs display the general strategy that I have used to subclone the 240-bp XbaI/AluI fragment (pCA1009, pCA1010; Appendix 2) and the 117-bp AluI fragment (pCA1007, pCA1012; Appendix 2) in Gemini 3 or Gemini 4. Both fragments are derived from the 5'-flanking region (indicated by a filled box) of the mouse pgk-1^b gene (Adra et al., 1987).


pCA BI7 pr Xba I/Taq I
520; G4
Alu I
Stu I
Alu I
Taq I
Pst I
Sph I
Hind III/Alu I
Ori
Eco RI
Xba I
Bam HI
Ava I
Sma I
Kpn I
Sac I
Alu I
Eco RI
Xmn I
Sca I
Pvu I
Mst I
Bgl I
AmpR
pCA 1009
pCA 1010
pCA 1007
pCA 1012
Alu I/Hinc II
Taq I
Pst I
Sph I
Alu I/Hinc II
Pst I
Sph I
Hind III
Xba I
Bam HI
Ava I
Sma I
Kpn I
Sac I
Eco RI
pCA BI7 Alu I/Alu I
190; G3
(Deleted CAAT box)

APPENDIX 2 Cont'd

The recombinant pCAB17prStuI/TaqI200-GEM-4. It contains a 200-bp StuI/AluI fragment derived from the 5'-flanking region of the mouse pgk-1^b allele (Adra et al., 1987). The construct pCAB17prBamHI/TaqI919.GEM-4 (Appendix 2) was digested with StuI and HindIII. The gragment carrying the 200-bp StuI/AluI fragment (indicated by a filled box) was then inserted into a HindII/HindIII-digested Gemini 4.

pCA BI7 pr Stu I / Taq I 200-GEM™-4

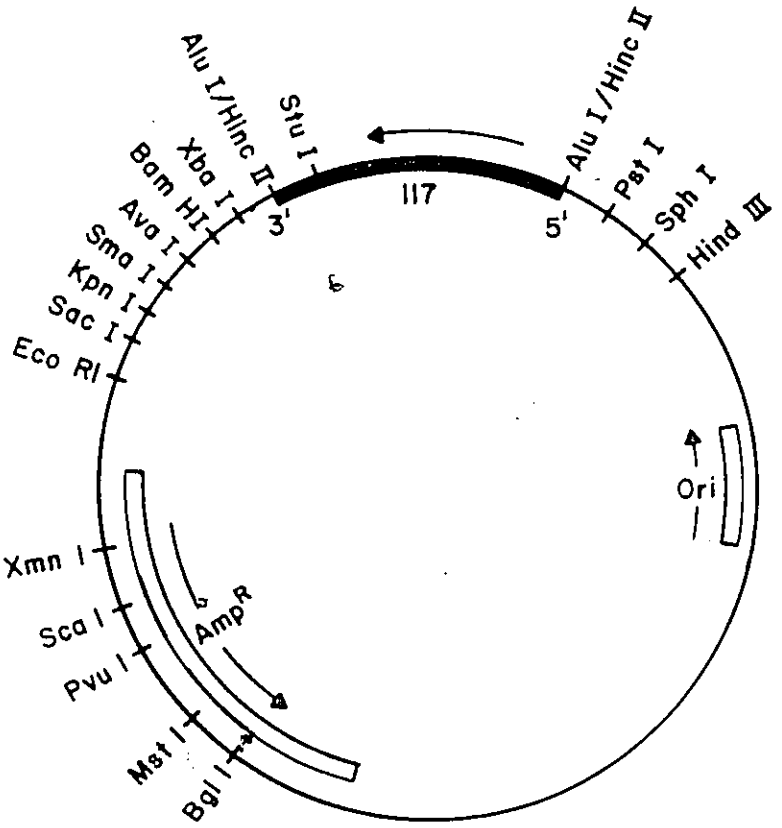


APPENDIX 2 Cont'd

The recombinant pCA1007. It contains a 117-bp AluI fragment derived from the 5'-flanking sequence of the mouse pgk-1^b allele (Adra et al., 1987). This construct was prepared by digesting pCAB17prXbaI/TaqI520;G4 (Appendix 2) with AluI. The 117-bp AluI fragment was excised from low melting agarose gel and subcloned into a HindII-digested Gemini 4 according to the strategy described in section 2.8.

pCA 1007

pCA B17 Alu I 117 G4 5'H ⟶ 3'E

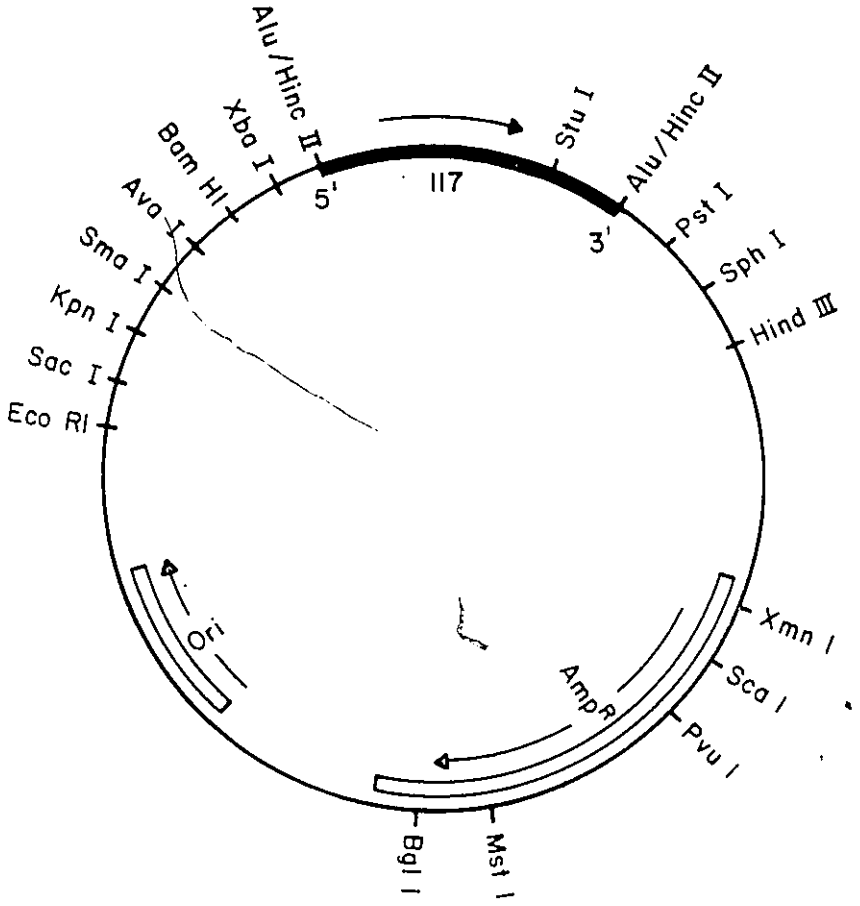


APPENDIX 2 Cont'd

The recombinant pCA1012. It contains a 117-bp AluI fragment derived from the 5'-flanking sequence of the pgk-1^b allele. This construct was constructed by digesting pCAB17prXbaI/TaqI520;G4 with AluI (Appendix 2). The 117-bp fragment (indicated by a filled box) was then subcloned into a HindII-digested Gemini 3.

pCA 1012

pCA B17 Alu I 117 G3 5'E⟶3'H

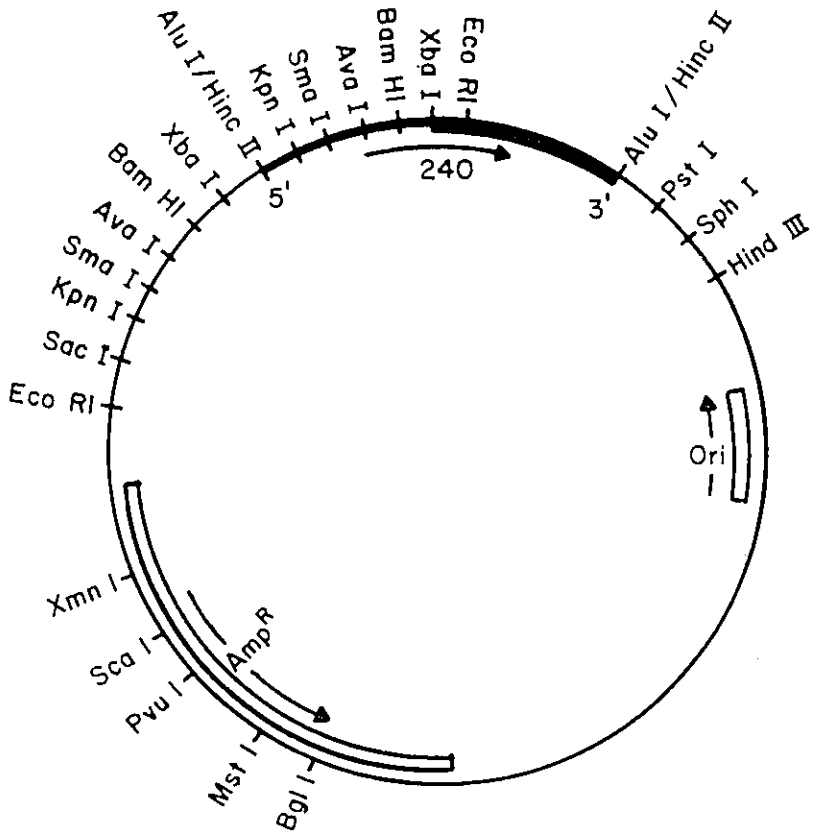


APPENDIX 2 Cont'd

The recombinant pCA1009. It contains a 240-bp XbaI/AluI fragment derived from the 5'-flanking sequence of the mouse pgk-1^b allele. This construct was prepared by digesting pCAB17prXbaI/TaqI520;Gr (Appendix 2) with AluI. An AluI fragment containing the XbaI/AluI fragment (indicated by a filled box) and a part of the multiple cloning site (indicated by a thick line) was then subcloned into a HindII-digested Gemini 4.

PCA 1009

pCA B17 Alu I 240 G4 5'E → 3'H

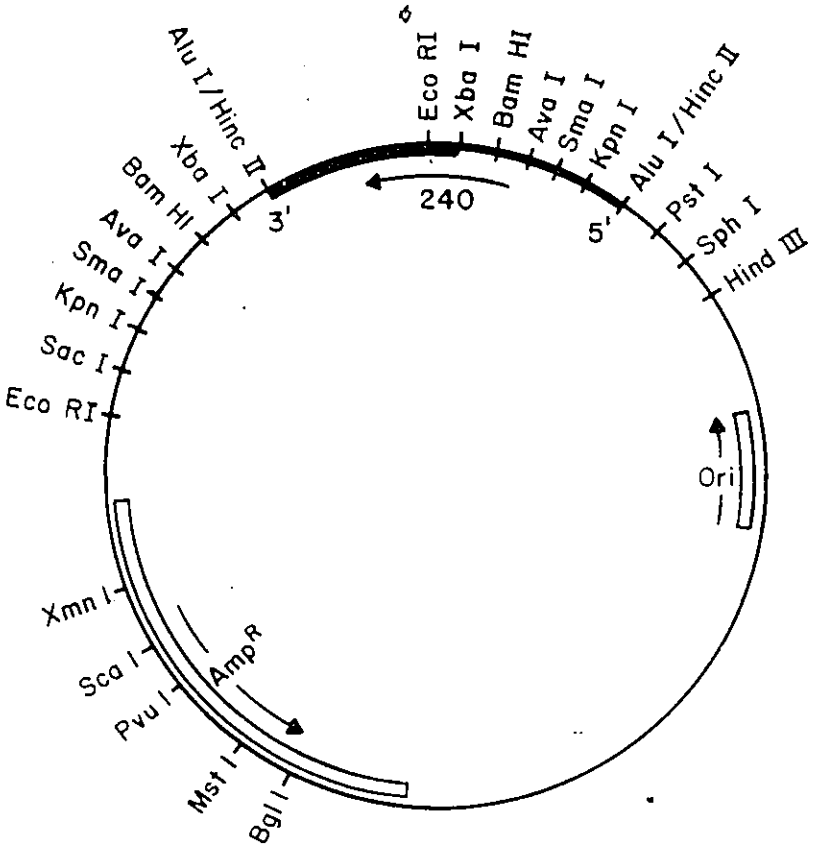


APPENDIX 2 Cont'd

The recombinant pCA1010. It contains the same insert as pCA1009 except that it is inserted in the reverse orientation.

pCA 1010

pCA B17 Alu I 240 G4 5'H→3'E

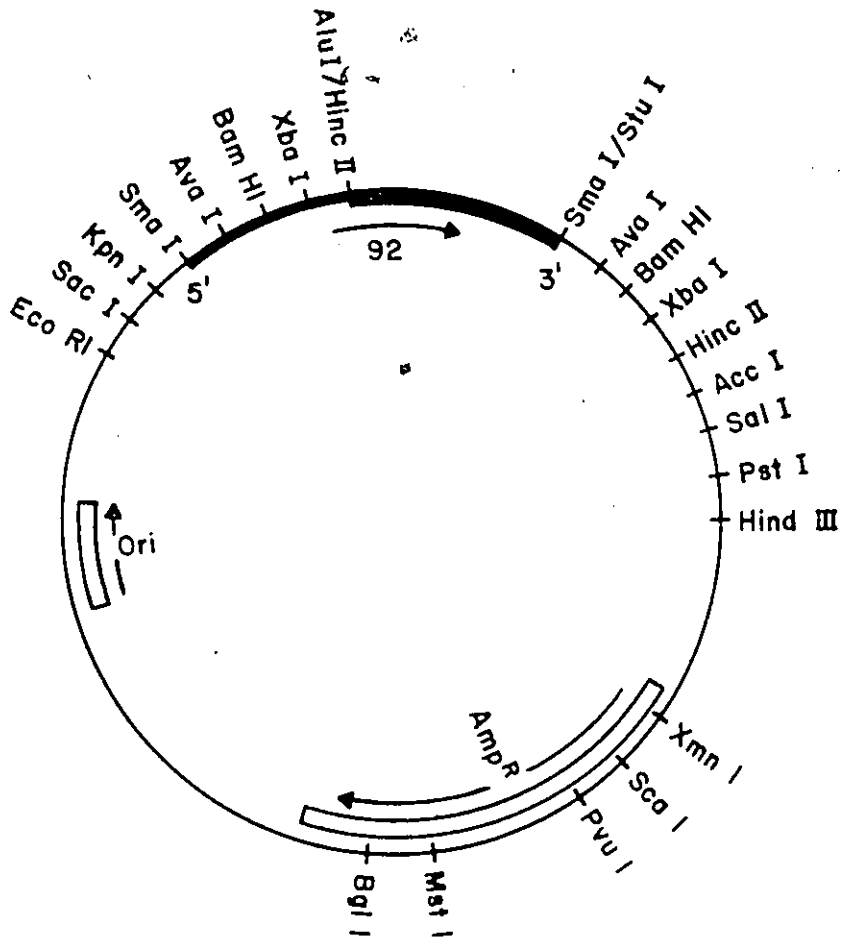


APPENDIX 2 Cont'd

The recombinant pCA1019. A 92-bp AluI-StuI fragment (indicated by a filled box) derived from the 5'-flanking region of the mouse pgk-1^b allele (Adra et al., 1987) was subcloned into the Gemini 3 vector.

pCA 1019

pCA BI7 Alu I-Stu I 92 G3 5'E⟶3'H

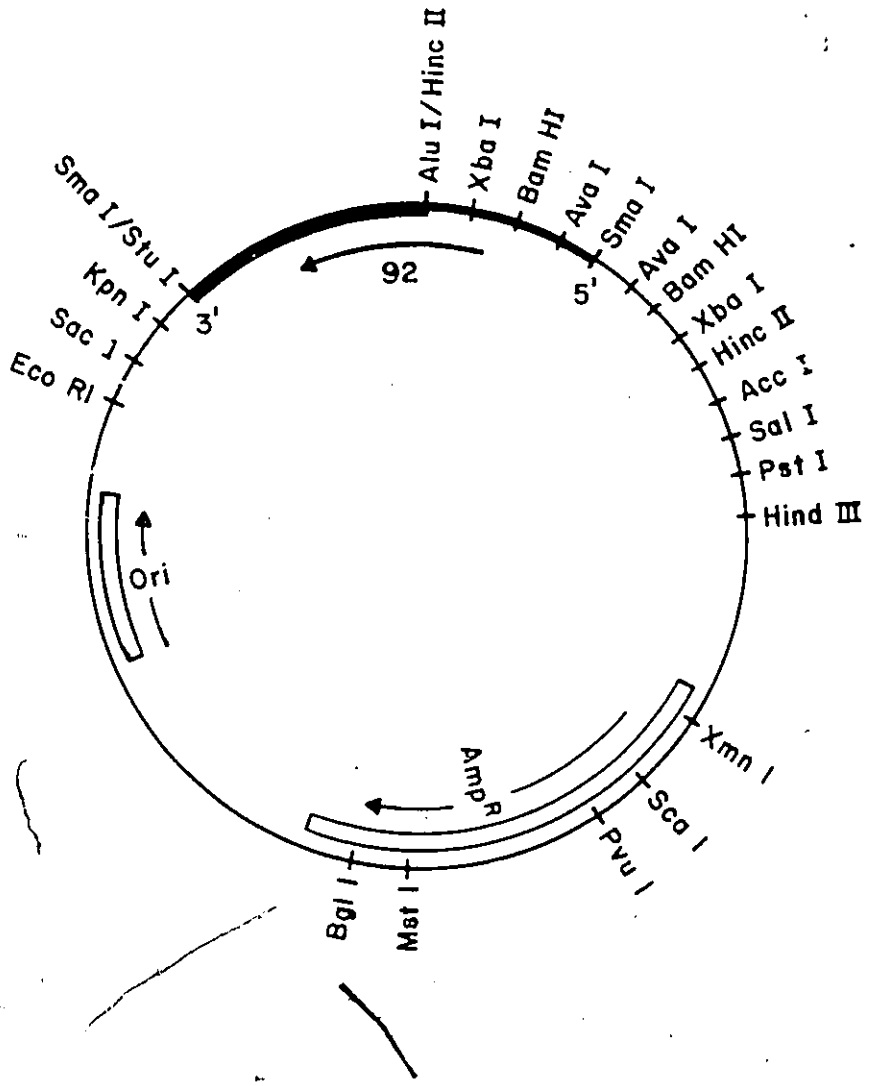


APPENDIX 2 Cont'd

The recombinant pCA1021. It contains the same insert (indicated by a filled box) as pCA1019 except that it is inserted in the reverse orientation.

pCA 1021

pCA B17 Alu I-Stu I 92 G-3 5'H ⟶ 3'E



APPENDIX 3

I have attempted to clone the mouse pgk-1 cDNA in collaboration with Dr. André Royale and Mrs. Jane Craig through the use of synthetic oligonucleotides. We have isolated a single clone which I characterized using the M13 cloning and sequencing system. Here I describe our work toward the cloning of the pgk-1 cDNA.

The use of oligodeoxynucleotides (14-mer) to isolate the pgk-1 cDNA

Two cDNA libraries have been constructed, by Dr. André Royale (Institut du Cancer à Montréal), in pBR322 from mRNA isolated from mouse salivary glands and mouse liver. Based on the known amino acid sequence, a mixture of 4 oligodeoxynucleotides (14-mer) (Fig. 1) was synthesized (Biologicals) so as to be complementary to a region of pgk-1 mRNA. The probing of the cDNA libraries was carried out in collaboration with Mrs. Jane Craig. The oligodeoxynucleotide was labelled at the 5' end with ^{32}P by polynucleotide kinase and gamma-^{32}P-ATP. The radioactive probe was used to screen the cDNA libraries. We isolated one clone, called pM-1, which specifically reacts with the oligodeoxynucleotide. The cDNA insert which is about 600 base pairs was excised from the vector by digestion with Pst-1, purified from other nucleic acids by agarose gel electrophoresis and by electroelution then cloned into the Pst-1 site of M13mp8. The ligated DNA was then mixed with the competent E. coli host cells and the cells were plated on media with IPTG and X gal. White plaques, that is, M13 infected cells that are lac$^-$, were isolated. To identify

phage clones containing the insert, labelled oligonucleotides were used as in situ hybridization probes against replica filters containing white and blue plaques. Several white plaques hybridized to the probes (Fig. 2). To know the orientation of the templates to be sequenced, Southern blotting of the double-stranded (RF) DNA (Fig. 3) and of the single-stranted DNA was performed (Fig. 4). The RF DNA of the two white plaques 14 and 40 hybridized to the labelled oligonucleotides whereas only the single-stranded viral form of white plaque 40 hybridized to the probes. Therefore the insert in the viral strand of the recombinant phage 40 is complementary to the insert in the viral strand of phage 14.

I prepared the two single-stranded DNA templates and I have done the sequencing (Sanger et al., 1977) of the templates using a "universal" primer (Heidecker et al., 1980; Messing et al., 1981). The sequencing (Fig. 5) of the cDNA established that the clone which reacted with the 14-mer was not made from pgk mRNA (Fig. 6, 7). To examine whether this cDNA insert is homologous to sequences located on the X chromosome, human DNA samples containing different copy numbers of the X chromosome were digested with EcoR1 and analyzed by Southern blot hybridization. None of the bands exhibited dosage of the hybridization signal with the number of X chromosomes.

The sequences of this cDNA insert and of the protein that can be translated from it are not homologous to those available in the DNA and protein data banks.

pgk-1 oligonucleotide mixture

protein N--phe-gln-trp-gln-ala---C

amino acid 343 344 345 346 347

mRNA 5'-UUPyGAPuUGGGAPuGC-3'

oligo mixture 5'-GCPyTCCCAPyTGGAA-3'

Fig. 1: pgk oligonucleotide sequence. This mixture consisted of 4 sequences. P_y means an equal mixture of T and C. P_u means an equal mixture of two deoxypurines A and G.

Fig. 2: In situ hybridization of white and blue plaques with the pgk oligonucleotides. 46 plaques were chosen. A replica filter was prepared with the plaques arranged in the grid pattern in A. B, the filter was hybridized with the ^{32}P-labelled oligonucleotides.

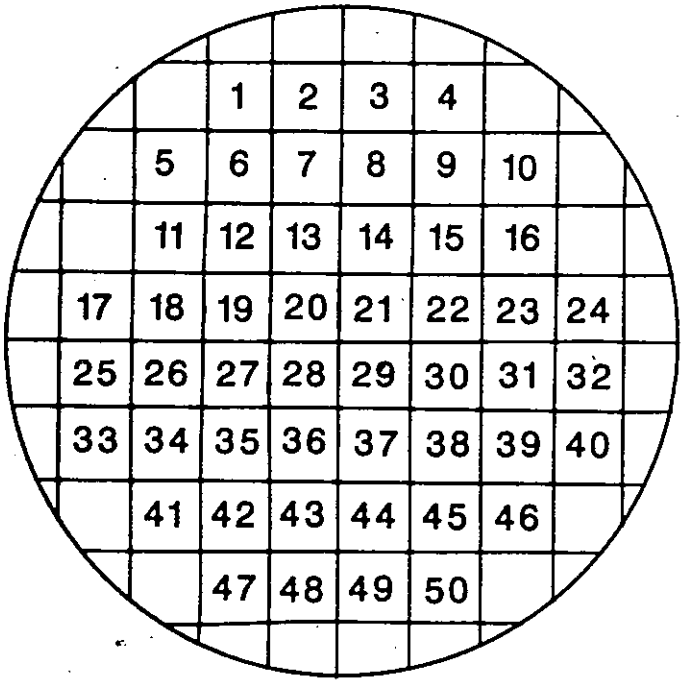

1 2 3 4
5 6 7 8 9 10
11 12 13 14 15 16
17 18 19 20 21 22 23 24
25 26 27 28 29 30 31 32
33 34 35 36 37 38 39 40
41 42 43 44 45 46
47 48 49 50

Fig. 3: Identification of recombinants by Southern blotting of RF DNA. The 600 bp insert was purified by electroelution and cloned into the PstI site of M13 mp8. Two blue plaques (bp) and 5 white plaques (wp) were selected for analysis. The DNA was probed with the ^{32}P-labelled oligonucleotides. Lanes: 1, wp #16; 2, wp #15; 3, wp #14; 4, bp #46; 5, wp #40; 6, bp #46, 7, wp #45; 8, bp #21; 9, pM-1 DNA digested with PstI.

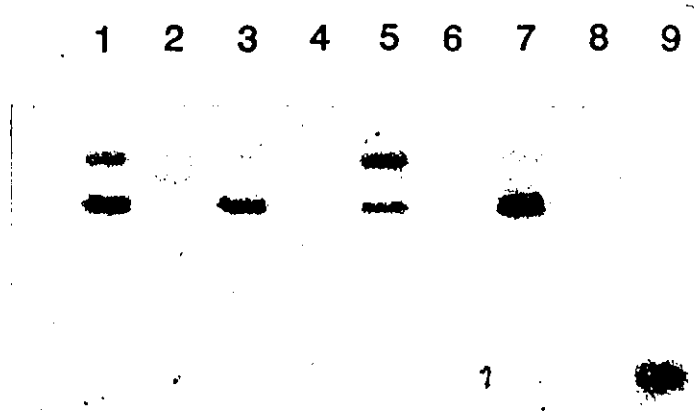

1 2 3 4 5 6 7 8 9

Fig. 4: Identification of recombinants by Southern blotting of the single-stranded viral form. Plasmid DNA was isolated from the clone which displayed the most stringent hybridization with the PGK oligonucleotides. Digestion with PstI revealed a 600 bp insert. The insert was purified by electroelution and cloned into the PstI site of M13 mp8. Three white plaques (wp) and one blue plaque (bp) were selected for analysis. The DNA was probed with ^{32}P-labelled oligonucleotides. Lanes: 1, wp #40; 2, bp #30; 3, wp #15; 4, wp #14.

1 2 3 4

Fig. 5: Sequencing gel showing part of the sequence of the inserts in the viral strands of phage 14 (lanes 1-4) and of phage 40 (lanes 5-8). The sequence was determined by the chain-termination technique (Sanger et al., 1977) using 8% urea-acrylamide gel (18 x 85 cm, 0.4 mm thick) in 1 x TBE. The oligo (G) tails flanking the inserts are shown. Lanes: 1 and 5 contained reaction A; 2 and 6 contained reaction T; 3 and 7 contained reaction G; 4 and 8 contained reaction C.

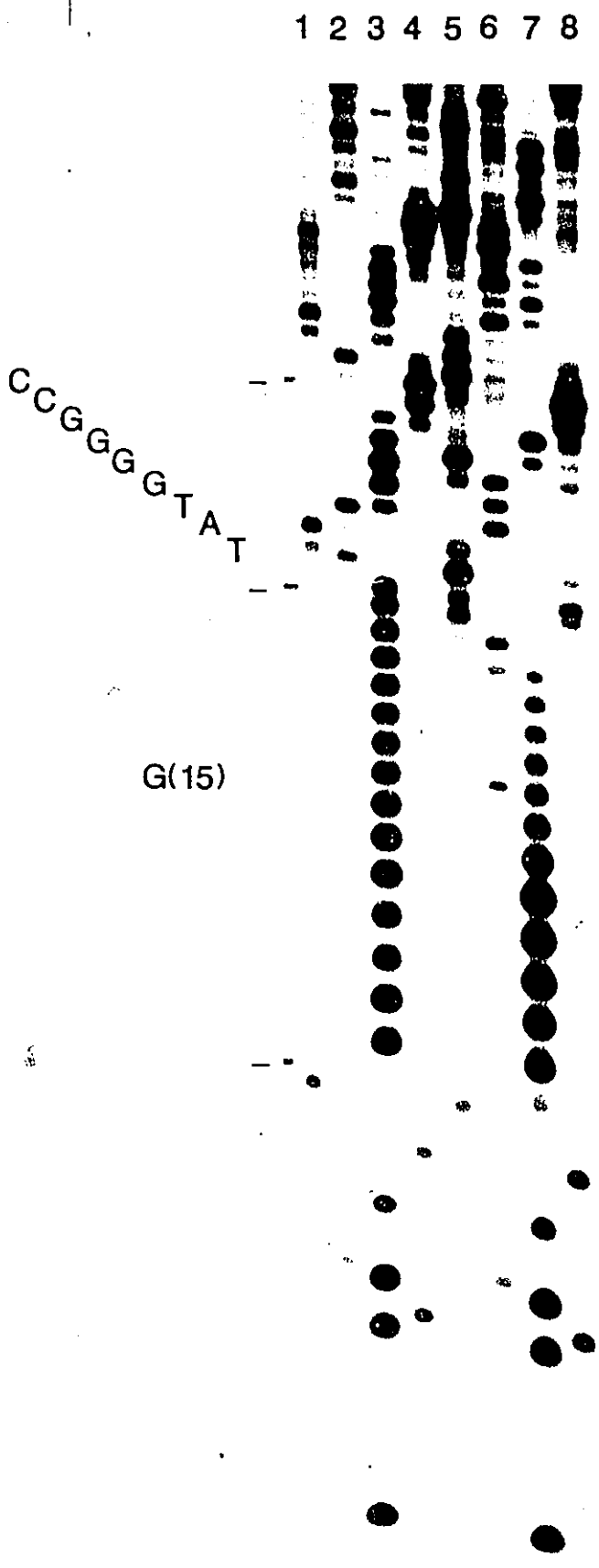

1 2 3 4 5 6 7 8
CCGGGGTAT
G(15)

Fig. 6: The clone which reacted with the 14-mer was not made from pgk mRNA. Nucleotide sequences of the 5' end of the insert in the viral strand of phage 40. The top line represents the insert sequence and the bottom line represents the human pgk cDNA sequence, where nucleotides identical to those in the insert are indicated by asterisks. Dashes represent gaps. The dashes were used to emphasize regions of sequence homology.

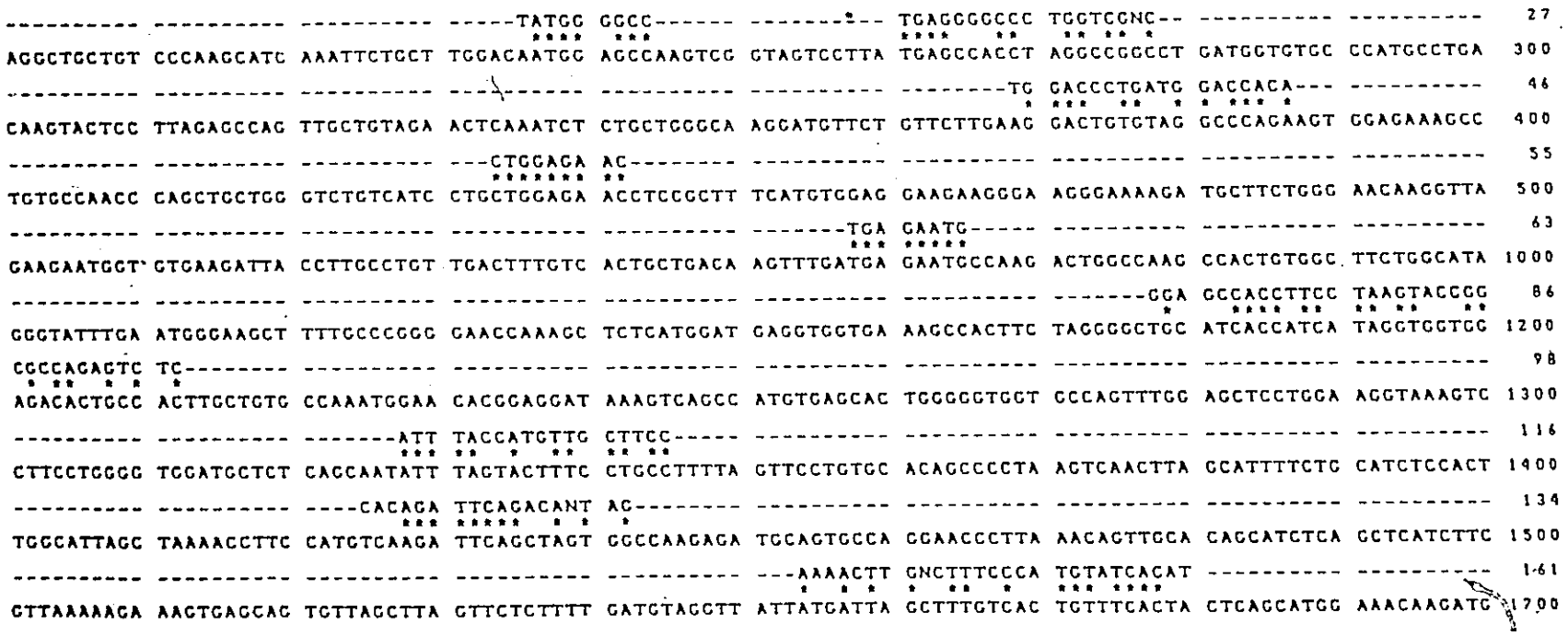
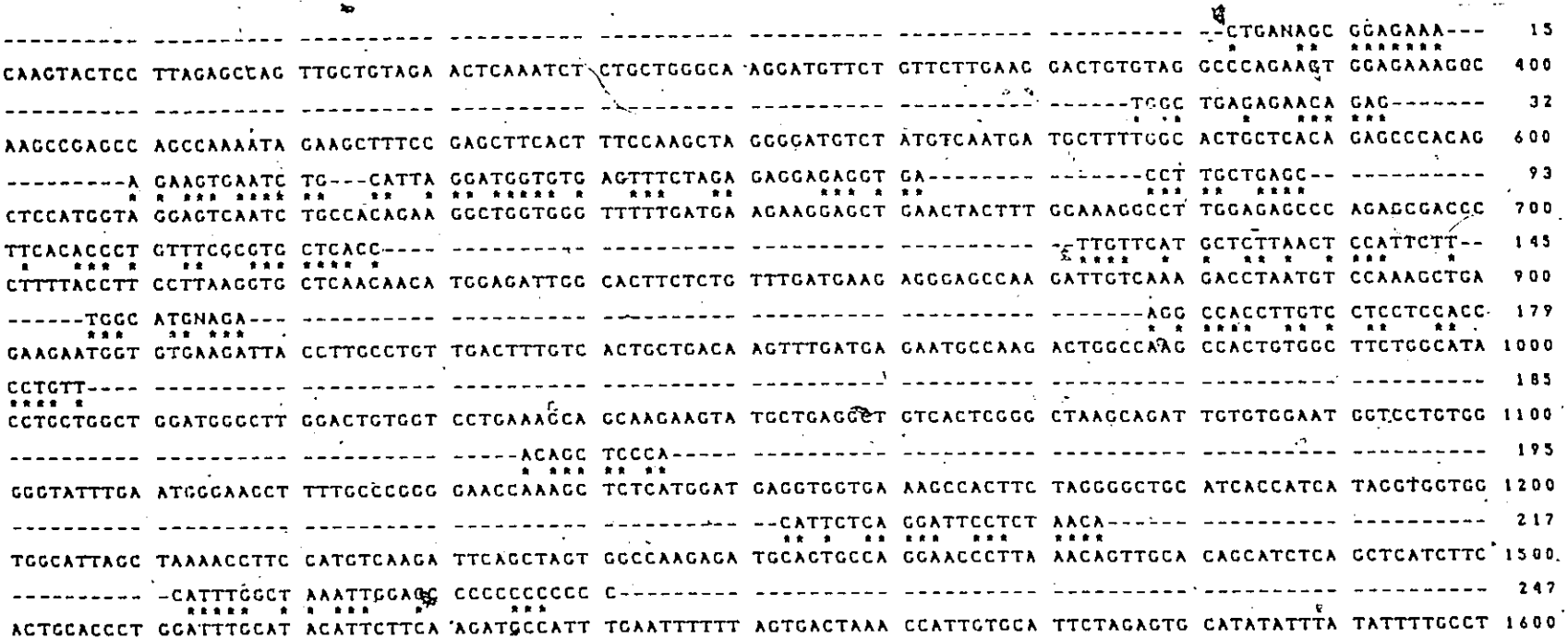


Fig. 7: The clone which reacted with the 14-mer was not made from pgk mRNA. Nucleotide sequences of the 3' end of the insert in the viral strand of phage 40. The top line represents the insert sequence and the bottom line represents the human pgk cDNA sequence, where nucleotides identical to those in the insert are indicated by asterisks. Dashes represent gaps. The dashes were used to emphasize regions of sequence homology.



REFERENCES

Adams, R., Burdon, R. and Fulton, J. (1983) Methylation of satellite DNA. Biochem. Biophys. Res. Commun. 113: 695-702.

Adra, C.N., Boer, P.H. and McBurney, M.W. (1987) Cloning and expression of the mouse pgk-1 gene and the nucleotide sequence of its promoter. Gene 60: 65-74.

Adra, C.N., Ellis, N.A. and McBurney, M.W. (1988) The family of mouse phosphoglycerate kinase genes and pseudogenes. Somat. Cell and Mol. Genet. 14: 69-81.

Auffray, C. and Rougeon, F. (1980) Purufication of mouse immunoglobulin heavy chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. 107: 303-314.

Baker, W.K. (1968) Position-effect variegation. Adv. Genet. 14: 133-169.

Baverstock, P.R., Adams, M., Polkinghorne, R.W. and Gelder, M. (1982) A sex-linked enzyme in birds - Z chromosome conservation but no dosage compensation. Nature 296: 763-766.

Benoist, C. and Chambon, P. (1981) *In vivo* sequence requirements of the SV40 early promoter region. Nature 290: 304-310.

Benson, W.D. and Davis, R.W. (1977) Screening of lambda gt recombinant clones by hybridization to single plaques *in situ*. Science 196: 180-182.

Berger, F.G., Szymanski, P., Read, E. and Watson, G. (1984) Androgen-regulated ornithine decarboxylase mRNAs of mouse kidney. J. Biol. Chem. 259: 7941-7946.

Bestor, T.H. and Ingram, V.M. (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity and mode of interaction with DNA. Proc. Natl. Acad. Sci. USA 80: 5559-5563.

Bird, A.P. (1986) CpG-rick islands and the function of DNA methylation. Nature 321: 209-213.

Bird, A.P., Taggart, M., Frommer, M., Miller, O.J. and Macleod, D. (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell 40: 91-99.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7: 1513-1523.

Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Transcription termination and 3' processing: the end is in site. Cell 41: 349-359.

Blake, C. (1983) Exons present from the beginning. Nature 306: 535-537.

Blattner, F.R., Williams, B.G., Blechl, A.E., Denniston-Thompson, K., Faber, H.E., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L. and Smithies, O. (1977) Charon phages: Safer derivatives of bacteriophage lambda for DNA cloning. Science 196: 161-169.

Blin, N. and Stafford, D.W. (1976) A general method for isolation of high molecular weight DNA from eukaryotes. Nucl. Acids Res. 3: 2303-2308.

Boer, P.H., Adra, C.N., Lau, Y.F. and McBurney, M.W. (1987) The testis-specific phosphoglycerate kinase gene, pgk-2, is a recruited retroposon. Mol. Cell. Biol. 7: 3107-3112.

Bolivar, F. and Backman, K. (1979) Plasmids of Escherichia coli as cloning vectors: Methods in Enzymology, Vol. 68, Wu, R. (ed.). Academic Press, New York and London, 245-267.

Boyer, H.W. and Roulland-Dussoix, D. (1969) A complementation analysis of the restriction and modification of DNA in Escherichia coli. J. Mol. Biol. 41: 459-472.

Brack, C. and Tonegawa, S. (1977) Variable and constant parts of the immunoglobin light chain gene of a mouse myeloma cell are 1250 nontranslated bases apart. Proc. Natl. Acad. Sci. USA 74: 5652-5656.

Breathnach, R. and Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. Ann. Rev. Biochem. 50: 349-383.

Broderick, T.P., Schaff, D.A., Bertino, A.M., Duah, M.K., Tischfield, J.A. and Sambrook, P.J. (1987) Comparative anatomy of the human APRT gene and enzyme: nucleotide sequence divergence and conservation of nonrandom CpG dinucleotide arrangement. Proc. Natl. Acad. Sci. USA 84: 3349-3353.

Brown, W.R.A. and Bird, A.P. (1986) Long-range restriction site mapping of mammalian genomic DNA. Nature 322: 477-483.

Bucher, T., Bender, W., Fundele, R., Hofner, H. and Linke, I. (1980) Quantitative evaluation of electrophoretic allo and isozyme patterns. FEBS Lett. 115: 319-325.

Burch, J.B.E. and Weintraub, H. (1983) Temporal order of chromatin structural change associated with activation of the major chicken vitellogenin gene. Cell 33: 65-76.

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.

Busslinger, M., Hurst, J. and Flavell, R. (1983) DNA methylation and the regulation of globin gene expression. Cell 34: 197-206.

Carle, G.F., Frank, M. and Olson M.V. (1986) Electrophoretic separations of large DNA molecules by periodic inversion of the electric field. Science 232: 65-68.

Cate, R.L., Chick, W. and Gilbert, W. (1983) Comparison of the methylation patterns of the two rat insulin genes. J. Biol. Chem. 258: 6645-6652.

Cattanach, B.M. (1970) Controlling elements in the mouse X-chromosome. III. Influence upon both parts of an X divided by rearrangement. Genet. Res. 16: 293-301.

Cattanach, B.M. (1975) Control of chromosome inactivation. Ann. Rev. Genet. 9: 1-18.

Cattanach, B.M., Perez, J.N. and Pollard, C.E. (1970) Controlling elements in the mouse X-chromosome. II. Location in the linkage map. Genet. Res. 15: 189-195.

Cattanach, B.M., Polalrd, C.E. and Perez, J.N. (1969) Controlling elements in the mouse X-chromosome. I. Interaction with the X-linked genes. Genet. Res. 14: 223-235.

Cattanach, B.M. and Isaacson, J.H. (1967) Controlling elements in the mouse X-chromosome. Genetics 57: 331-346,

Cattanach, B.M., and Williams, C.E. (1972) Evidence of nonrandom X-chromosome activity in the mouse. Genet. Res. 19: 229-240.

Cattanach, B.M., Bucher, T.H. and Andrews, S.J. (1983) pgk-1 expression in pgk-1^a/pgk-1^b female mice associated with Xce allele substitution. Genet. Res. 41: 312.

Cattanach, B.M. and Papworth, P. (1981) Controlling elements in the mouse. V. linkage tests with X-linked genes. Genet. Res. 38: 57-70.

Chandra, H.S. and Brown, S.W. (1975) Chromosome imprinting and the mammalian X chromosome. Nature 253: 165-168.

Chapman, V., Forrester, C., Sanford, J., Hastie, N. and Rossant, J. (1984) Cell lineage-specific undermethylation of mouse repetitive DNA. Nature 307: 284-286.

Chapman, V.M., Kratzer, P.G., Siracusa, L.O., Quarantillo, B.A., Evans, R. and Liskay, R.M. (1982) Evidence for DNA modification in the maintenance of X-chromosome inactivation of adult mouse tissues. Proc. Natl. Acad. Sci. USA 79: 5357-5361.

Chen, C.Y. and Hitzeman, R.A. (1987) Human, yeast and hybrid 3-phosphoglycerate kinase gene expression in yeast. Nuc. Acids Res. 15: 643-660.

Church, G. and Gilbert, W. (1984) Genomic sequencing. Proc. Natl. Acad. Sci. USA 81: 1991-1995.

Clark, B.D., Collins, K.L., Gandy, M.S., Webb, A.C. and Auron, P.E. (1986) Genomic sequence for human prointerleukin 1 beta: possible evolution from a reverse transcribed prointerleukin 1 alpha gene. Nucl. Acids. Res. 14: 7897-7914.

Cock, A.G. (1964) Dosage compensation and sex chromatin in nonmammals. Genet. Res. 5: 354-365.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon P. (1980) Promoter sequences of eucaryotic protein-coding genes. Science 209: 1406-1413.

Cornish-Bowden, A. (1985) Are introns structural elements or evolutionary debris? Nature 313: 434-435.

Cullen, D.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 J. 5: 2223-2229.

Doskocil, J. and Sorm, F. (1962) Distribution of 5-methylcytosine in pyrimidine sequences of DNA. Biochem. Biophys. Acta 55: 953-959.

Dynan, W.S. and Tjian, R. (1985) Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature 316: 774-778.

Echols, H. (1986) Multiple DNA-protein interactions governing high-precision DNA transactions. Science 233: 1050-1056.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.-H., Midgett, R.M., Kuo, K.C., McCune, R.A. and Gehrke, C. (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Res. 10: 2709-2721.

Eicher, E.M. (1970) X-autosome translocations in the mouse: total inactivation versus partial inactivation of the X-chromosome. Adv. Genet. 15: 175-259.

Eicher, E.M., Cherry, M. and Flaherty, L. (1978) Autosomal phosphoglycerate kinase linked to mouse major histocompatibility complex. Mol. Gen. Genet. 158: 225-228.

Erickson, R.P., Michelson, A.M., Rosenberg, M.P., Sanchez, E. and Orkin, S.H. (1985) Post-meiotic transcription of phosphoglycerate kinase-2 in mouse testes. Biosci. Rep. 5: 1087-1091.

Esfradiatis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spiritz, R.A., De Riel, J.K., Forget, B.G., Slighton, L., Blechl, A.E., Smithies, O., Barralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21: 653-668.

Everett, R.D., Baty, D. and Chambon, P. (1983) The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter. Nucl. Acids Res. 11: 2447-2464.

Falconer, D.S. and Isaacson, J.H. (1972) Sex-linked variegation modified by selection in brindled mice. Genet. Res. 20: 291-316.

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

Fisher, J.H., Gusella, J.F. and Scoggin, C.H. (1984) Molecular hybridization under conditions of high stringency permits cloned DNA segments containing reiterated DNA sequences to be assigned to specific chromosomal locations. Proc. Natl. Acad. Sci. USA 81: 520-524.

Flejter, W.L., Van Dyke, D.L. and Weiss, L. (1986) Location of the X inactivation center in primates and other mammals. Hum. Genet. 74: 63-66.

Flemington, E., Bradshaw, H.D., traina-Dorge, V., Slagel, V. and Deininger, P.L. (1987) Sequence, structure and promoter characterization of the human thymidine kinase gene. Gene 52: 267-277.

Fradin, A., Manley, J. and Prives, C. (1982) Methylation of simian virus 40 HpaII site affects late, but not early, viral gene expression. Proc. Natl. Acad. Sci. USA 79: 5142-5146.

Fried, M. and Crothers, D.M. (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucl. Acids Res. 9: 6505-6525.

Frischauf, A.M., Lehrach, H., Poustka, A. and Murray, N. (1983) Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. 170: 827-842.

Fuscoe, J.C., Fenwick, R.G., Jr., Ledbetter, D.H. and Caskey, C.T. (1983) Deletion and amplification of the HPRT locus in Chinese hamster cells. Mol. Cell. Biol. 3: 1086-1096.

Galas, D.J. and Schmitz, A. (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucl. Acids Res. 5: 3157-3170.

Gartler, S.M., Andina, R. and Gant, N. (1975) Ontogeny of X-chromosome inactivation in the female germ line. Exp. Cell Res. 91: 454-457.

Gartler, S.M., Rivest, M. and Cole, R.E. (1980) Cytological evidence for an inactive X chromosome in murine oogonia. Cytogenet. Cell Genet. 28: 203-207.

Gartler, S.F. and Riggs, A.D. (1983) Mammalian X-chromosome inactivation. Ann. Rev. Genet. 17: 155-190.

Gerber-Huber, S., May, F.E.B., Westley, B.R., Felber, B.K., Hosbach, H.A., Andres, A.-C. and Ryffel, G.U. (1983) Cell 33: 43-51.

Gerlach, W.L. and Bedbrook, J.R. (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res. 7: 1869-1885.

Gilbert, W. (1985) Genes-in-pieces revisited. Science 228: 823-824.

Gitshier, J., Wood, W.I., Goralka, T.M., Wion, K.L., Chen, E.Y., Eaton, D.H., Vehar, G.A., Capon, D.J. and Lawn, R.M. (1984) Characterization of the human factor VIII gene. Nature 312: 326-330.

Gold, B., Fryimato, H., Kramer, J.M., Erickson, R.P. and Hecht, N.B. (1983) Haploid accumulation and translational control of phosphoglycerate kinase-2 messenger RNA during mouse spermatogenesis. Dev. Biol. 98: 392-399.

Goldberg, M. (1979) Sequence Analysis of Drosophila Histone Genes. Ph.D. Thesis, Stanford University, Stanford, CA.

Goodman, H.M., Olson, M.V. and Hall, B.D. (1977) Nucleotide sequence of a mutant eukaryotic gene: the yeast tyrosine-inserting ochre suppressor SUP4-o. Proc. Natl. Acad. Sci. USA 74: 5453-5457.

Graham, F.L. and Van der Eb, A.J. (1973) A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52: 456-467.

Grahn, D., Lea, R.A. and Hulesch, J. (1970) Linkage analysis of a presumed X-inactivation controlling element. Mouse News Lett. 42: 16.

Grainger, R.M., Hazard-Leonards, R.M., Samaha, F., Hougan, L.M., Lesk, M.R. and Thomsen, G.H. (1983) Is hypomethylation linked to activation of δ-crystallin genes during lens development. Nature 306: 88-90.

Graves, J.A.M. (1982) 5-azacytidine-induced re-expression of alleles on the inactive X chromosome in a hybrid mouse cell line. Exp. Cell Res. 141: 99-105.

Gruenbaum, Y., Cedar, H. and Razin, A. (1982) Substrate and sequence specificity of a eukaryotic DNA methylase. Nature 295: 620-622.

Gubler, U. and Hoffman, B.J. (1983) A simple and very efficient method for generating cDNA libraries. Gene 25: 263-269.

Gusella, J.F., Jones, C., Kao, F.-T., Housman, D. and Puck, T.T. (1982) Genetic fine-structure mapping in human chromosome 11 by use of repetitive DNA sequences. Proc. Natl. Acad. Sci. USA 79: 7804-7808.

Hanahan, D. and Meselson, M. (1980) Plasmid screening at high colony density. Gene 10: 63-67.

Hall, C.V., Jacob, P.E., Gordon, M.R. and Lee, F.L. (1983) Expression and regulation of Escherichia coli lacZ gene fusions in mammalian cells. J. Mol. Appl. Genet. 2: 101-109.

Hecht, N.B. (1987) Regulation of gene expression during mammalian spermatogenesis. Experimental Approaches to Mammalian Embryonic Development, J. Rossant and R.A. Peterson (eds), Cambridge University Press, London.

Hendrix, R.W., Roberts, J.W., Stahl, F. and Weisberg, R.A. (1983) Lambda II. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.

Hohn, B. and Collins, J. (1980) A small cosmid for efficient cloning of large DNA fragments. Gene 11: 291-298.

Holliday, R. and Pugh, J.E. (1975) DNA modification mechanisms and gene activity during development. Science 187: 226-232.

Huynh, T.V., Young, R.A. and Davis, R.W. (1985) Construction and screening cDNA libraries in gamma gt 10 and gamma gt 11. In DNA Clonigh Techniques: A practical Approach, D.M. Glover (ed), Vol. 1, IRL Press, Oxford, pp. 1-48.

Ingolia, D.E., Al-Ubaidi, M.R., Yeung, C.Y., Bigo, H.A., Wright, D.A. and Kellems, R.E. (1986) Molecular cloning of the murine adenosine deaminase gene from a genetically enriched source: identification and characterization of the promoter region. Mol. Cell. Biol. 6: 4458-4466.

Jähner, D., Stuhlmann, H., Stewart, C.L., Harbers, K., Löhler, J., Simon, I. and Jaenisch, R. (1982) De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature 298: 623-628.

Jeffreys, A.J. and Flavell, R.A. (1977) The rabbit beta-globin gene contains a large insert in the coding sequence. Cell 12: 1097-1108.

Jeffreys, A.J. and Harris, S. (1982) Processes of gene duplication. Nature 296: 9-10.

Johnson, M.S. and Turner, J.R.G. (1979) Absence of dosage compensation for a sex-linked enzyme in butterflies (Heliconius). Heredity 43: 71-77.

Johnston, P.G. and Cattanach, B.M. (1981) Controlling elements in the mouse. IV. Evidence of non-random X-inactivation. Genet. Res. 37: 151-160.

Jones, K.A., Yamamoto, K.R. and Tjian, R. (1985) Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell 42: 559-572.

Joyner, A.L., Kornberg, T., Coleman, K.G., Cox, D.R. and Martin, G.R. (1985) Expression during embryogenesis of a mouse gene with sequence homology to the Drosophila engrailed gene. Cell 43: 29-37.

Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) Promoter-specific activation of RNA polymerase II transcription. Spl. Trends Biochem. Sci. 11: 20-23.

Kaiser, K. and Murray, N.E. (1985) The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries. DNA Cloning: A Practical Approach, D.M. Glover (ed) Vol. 1, IRL Press, Oxford, pp. 49-78.

Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980) Novel bacteriophage lambda cloning vector. Proc. Natl. Acad. Sci. USA 77: 5169-5176.

Karn, J., Brenner, S. and Barnett, L. (1983) New bacteriophage lambda vectors with positive selection for cloned inserts. Methods in Enzymology 101: 3-19.

Kaslow, D.C. and Migeon, B.R. (1987) DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: Evidence for multistep maintenance of mammalian X dosage compensation. Proc. Natl. Acad. Sci. USA 84: 6210-6214.

Keitges, E.A. and Palmer, C.G. (1986) Analysis of spreading of inactivation in eight X autosome translocations utilizing the high resolution RBG technique. Hum. Genet. 72: 231-236.

Keith, D.H., Singer-sam, J. and Riggs, A.D. (1986) Active X-chromosome DNA is unmethylated at eight CCGG sites clustered in guanine-plus- cytosine-rich island at the 5' end of the gene for phosphoglycerate kinase. Mol. Cell. Biol. 6: 4122-4125.

Koenig, M., Hoffman, E.P., Bertelson, C.J., Monaco, A.P., Feener, C. and Kunkel, L.M. (1987) Complete cloning of the Duchesne Muscular Dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 50: 509-517.

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

Korneluk, R.G., Quan, F. and Gravel, R.A. (1985) Rapid and reliable dideoxy sequencing of double-stranded DNA. Gene 40: 317-323.

Kruczek, I. and Doerfler, W. (1983) Expression of the chloramphenical acetyltransferase gene in mammalian cells under the control of adenovirus type 12 promoters: effect of promoter methylation on gene expression. Proc. Natl. Acad. Sci. USA 80: 7586-7590.

Krzanowska, H., and Wabik, B. (1971) Selection for expression of sex-linked gene Ms (Mosaic) in heterozygous mice. Genet. Pol. 12: 537-544.

Kuo, M.T., Mandel, J.L. and Chambon, P. (1979) DNA methylation: Correlation with DNaseI sensitivity of chicken ovalbumin and conalbumin chromatin. Nucleic Acids Res. 7: 2105-2113.

Lee, C.-Y., Niesel, D., Pegoraro, B. and Erickson, R.P. (1980) Immunological and structural relatedness of isozymes and genetic variants of 3-phosphoglycerate kinase from the mouse. J. Biol. Chem. 255: 2590-2595.

Lester, S.C., Korn, N.J. and DeMars, R. (1982) Derepression of genes on the human X-chromosome: evidence for differences in locus-specific rates of derepression and rates of transfer of active and inactive genes after DNA-mediated transformation. Somatic Cell Genet. 8: 265-284.

Lewis, E.B. (1950) The phenomenon of position effect. Adv. Genet. 3: 73-115.

Lifschytz, E. and Lindsley, D.L. (1974) Sex chromosome activation during spermatogenesis. Genetics 78: 323-331.

Lindsay, S., Monk, M., Holliday, R., Huschtscha, L., Davies, K.E., Riggs, A.D. and Flavell, R.A. (1985) Differences in methylation on the active and inactive human X-chromosome. Ann. Hum. Genet. 49: 115-127.

Liskay, R.M. and Evans, R.J. (1980) Inactive X chromosome DNA does not function in DNA-mediated cell transformation for the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA 77: 4895-4898.

Lock, L.F., Melton, D.W., Caskey, C.T. and Martin, G.R. (1986) Methylation of the mouse hprt gene differs on the active and inactive X chromosomes. Mol. Cell. Biol. 6: 914-924.

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: 36-46.

Loenen, W.A.M. and Brammar, W.J. (1980) A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene 20: 249-259.

Lyon, M.F. (1966) Lack of evidence that inactivation of the mouse X-chromosome is incomplete. Genet. Res. 8: 197-203.

Lyon, M.F., Searle, A.G., Ford, C.E. and Ohno, S. (1964) A mouse translocation suppressing sex-linked variegation. Cytogenetics 3: 306-323.

Maeda, N. and Smithies, O. (1986) The evolution of multigene families: human haptoglobin genes. Ann. Rev. Genet. 20: 81-108.

Mandel, J.L. and Chambon, P. (1979) DNA methylation: organ specific variations in the methylation pattern within and around ovalbumin and other chicken genes. Nucleic Acids Res. 7: 2081-2103.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbour, New York: Cold Spring Harbour Laboratory.

Martin, G.R., Epstein, C.J., Travis, B., Tucker, G., Yatsiv, S., Martin, D.M., Clift, S. and Cohen, S. (1978) X-chromosome inactivation during differentiation of female teratocarcinoma cells in vitro. Nature 271: 329-333.

Martini, G., Toniolo, D., Vulliamy, T., Luzzato, L., Dono, R., Viglietto, G., Paonessa, G., D'Urso, M. and Persico, M.G. (1986) Structural analysis of the X-linked gene encoding human glucose 6-phosphate dehydrogenase. EMBO J. 5: 1849-1855.

Mattei, M.G., Mattei, J.F., Ayme, S. and Giraud, F. (1982) X-autosome translocations: cytogenetic characteristics and their consequences. Hum. Genet. 61: 295-309.

McBurney, M.W. and Strutt, B.J. (1980) Genetic activity of X chromosomes in pluripotent female teratocarcinoma cells and their differentiated progeny. Cell 21: 357-364.

McCarrey, J.R. and Thomas, K. (1987) Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. Nature 326: 501-505.

McKnight, S. and Tjian, R. (1986) Transcriptional selectivity of viral genes in mammalian cells. Cell 46: 795-805.

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. Genet. Res. 41: 69-83.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984a) Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nuc. Acids Res. 12: 7035-7055.

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

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

Messing, J. (1983) New M13 vectors for cloning. Methods in Enzymology, R. Wu, L. Grossman, K. Moldave (eds) Vol. 101, Academic Press New York and London, pp. 20-78.

Michelson, A.M., Blake, C.F.F., Evans, S.T. and Orkin, S.H. (1985) Structure of the human phosphoglycerate kinase gene and the intron-mediated evolution and dispersal of the nucleotide-binding domain. Proc. Natl. Acad. Sci. USA 82: 6965-6969.

Michelson, A.M., Markham, A.F. and Orkin, S.H. (1983) Isolation and DNA sequence of a full-length cDNA clone for human X chromosome-encoded phosphoglycerate kinase. Proc. Natl. Acad. Sci. USA 80: 472-476.

Michelson, A.M., Bruns, G.A.P., Morton, C.C. and Orkin, S.H. (1985) The human phosphoglycerate kinase multigene family. HLA-associated sequences and an X-linked locus containing a

processed pseudogene and its functional counterpart. J. Biol. Chem. 260: 6982-6992.

Migeon, B.R., Sprenkle, J.A., and Do, T.T. (1981) Stability of the "two active X" phenotype in triploid somatic cells. Cell 18: 637-641.

Mohandas, T., Sparkes, R.S. and Shapiro, L.J. (1981) Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science 211: 393-396.

Mohandas, T., Sparkes, R.S. and Shapiro, L.J. (1982) Genetic evidence for the inactivation of a human autosomal locus attached to an inactive X chromosome. Am. J. Hum. Genet. 34: 811-817.

Monesi, V. (1971) Chromosome activities during meiosis and spermiogenesis. J. Reprod. Fert. 13: 1-14.

Monk, M. (1986) Methylation and the inactive X. Bio Essays 4: 204-208.

Mori, N., Singer-Sam, J. and Riggs, A.D. (1986a) Evolutionary conservation of the substrate-binding cleft of phosphoglycerate kinases. FEBS Lett. 204: 313-317.

Mori, N., Singer-Sam, J., Lee. C.-Y. and Riggs, A.D. (1986b) The nucleotide sequence of a cDNA clone containing the entire coding region for the mouse X-chromosome linked phosphoglycerate kinase. Gene 45: 275-280.

Mulligan, R.C. and Berg, P. (1980) Expression of a bacterial gene in mammalian cells. Science 209: 1422-1427.

Nishioka, Y., Leder, A. and Leder, P. (1980) An unusual alpha globin-like gene that has cleanly lost both globin intervening sequences. Proc. Natl. Acad. Sci. USA 77: 2806-2809.

Myers, R.M., Rio, D.C., Robbins, A.K. and Tjian, R. (1981) SV40 gene expression modulated by the cooperative binding of T antigen to DNA. Cell 25: 373-384.

Norrander, J., Kempe, T. and Messing, J. (1983) Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26: 101-106.

Ogden, J.E., Stanway, C., Kim, S., Mellor, J., Kingsman, A.J. and Kingsman, S.M. (1986) Efficient expression of the Saccharomyces cerevisiae PGK gene depends on an upstream activation sequence but does not require TATA sequences. Mol. Cell. Biol. 6: 4335-4343.

Ohno, S. (1969) Evolution of sex chromosomes in mammals. Ann. Rev. Genet. 3: 495-524.

Ohno, S., Christian, L., Attardi, B.J. and Kan, J. (1973) Modification of expression of the testicular feminization (Tfm) gene of the mouse by a "controlling element" gene. Nature London New Biol. 245: 92-93.

Pages, M. and Roizes, G. (1982) Tissue specificity and organization of CpG methylation in calf satellite DNAI. Nucleic Acids Res. 10: 565-576.

Palmiter, R.D., Chen, H.Y. and Brinster, R.L. (1982) Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. Cell 29: 701-710.

Patel, P.I., Framson, P.E., Caskey, C.T. and Chinault, C. (1986) Fine structure of the human hypoxanthine phosphoribosyltransferase gene. Mol. Cell. Biol. 6: 393-403.

Paterno, G.D., Adra, C.N. and McBurney, M.W. (1985) X chromosome reactivation in mouse embryonal carcinoma cells. Mol. Cell. Biol. 5: 2705-2712.

Paterno, G.D. and McBurney, M.W. (1985) X chromosome inactivation during induced differentiation of a female mouse embryonal carcinoma cell line. J. Cell Sci. 75: 149-163.

Perry, R. (1983) Consequence of myc invasion of immunoglobulin loci: facts and speculation. Cell 33: 647-649.

Pfeifer, G.P., Grunwald, S., Boehm, T.L.J. and Drahovsky, D. (1983) Isolation and characterization of DNA cytosine 5-methyltransferase from human placenta. Biochim. Biophys. Acta 740: 323-330.

Piechaczyk, M., Blanchard, J.M., Riaad-El Sabouty, S., Dani, C., Marty, L. and Jenateur, P. (1984) Unusual abundance of vertebrate 3-phosphate dehydrogenase pseudogenes. Nature 312: 469-471.

Poustka, A., Pohl, T.M., Barlow, D.P., Frischauf, A.-M. and Lehrach, H. (1987) Construction and use of human chromosome jumping libraries from NotI-digested DNA. Nature 325: 353-355.

Ptashne, M. (1986) Gene regulation by proteins acting nearby and at a distance. Nature 322: 697-701.

Queen, C. and Korn, L.J. (1983) A comprehensive sequence analysis program for the IBM personal computer. Nucl. Acids Res. 12: 581-599.

Rastan, S. (1982) Primary non-random X-inactivation caused by controlling elements in the mouse demonstrated at the cellular level. Genet. Res. 40: 139-147.

Rastan, S. and Robertson, E.J. (1985) X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. J. Embryol. Exp. Morph. 90: 379-388.

Razin, A. and Riggs, A.D. (1980) DNA methylation and gene function. Science 210: 604-610.

Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G., Brown, M.S., Goldstein. J.L. and Luskey, K.L. (1984) HMGCoA reductase: a negatively regulated gene with unusual promoter and 5' untranslated regions. Cell 38: 275-285.

Richards-Smith, B. and Elliott, R. (1984) Mapping of a family of repeated sequences in the mouse genome. Mouse News Letter 71: 46-47.

Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. 113: 237-251.

Riggs, A.D. (1975) X inactivation, differentiation and DNA methylation. Cytogenet. Cell Genet. 14: 9-25.

Riggs, A.D. and Jones, P.A. (1983) 5-methylcytosine, gene regulation, and cancer. Adv. Cancer Res. 40: 1-30.

Rogers, J.H. (1983) Retroposons defined. Nature 301: 460.

Rogers, J.H. (1985) The origin and evolution of retroposons. Int. Rev. Cytol. 93: 187-279.

Russell, L.B. (1971) Attempts to demonstrate different inactivating states for normal mouse X chromosomes. Genetics 68: 55-56.

Russell, L.B. and Cacheiro, N.L.A. (1978) The use of X-autosome translocation in the study of X-inactivation pathways and nonrandomness. In Genetics Mosaics and Chimeras in Mammals, L.B. Russell (ed), New York, Plenum, pp. 393-416.

Sager, R. and Kitchin, R. (1975) Selective silencing of eukaryotic DNA. Science 189: 426-433.

Salomon, R. and Kaye, A.M. (1970) Methylation of mouse DNA in vivo: Di and trypyrimidine sequences containing 5-methylcytosine. Biochem. Biophys. Acta 204: 340.

Sanford, J.P., Chapman, V.M. and Rossant, J. (1985) DNA methylation in extraembryonic lineages of mammals. Trends in Genetics 1: 89-93.

Sanford, J., Forrester, L., Chapman, V., Chandley, A. and Hastie, N. (1984) Methylation patterns of repetitive DNA sequences in germ cells of Mus musculus. Nuc. Acids Res. 12: 2823-2836.

Sanger, F., Nicklen, S. and Coulson, A. (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Sano, H., Noguchi, H and Sager, R. (1983) Characterization of DNA methyltransferase from bovine thymus cells. Eur. J. Biochem. 135: 181-185.

Schleif, R. (1987) Why should DNA loop. Nature 327: 369-370.

Schwartz, D.C. and Cantor, C.R. (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient electrophoresis. Cell 37: 67-75.

Shen, C.J. and Maniatis, T. (1980) Tissue-specific DNA methylation in a cluster of rabbit β-like globin genes. Proc. Natl. Acad. Sci. USA 77: 6634-6638.

Simon, D., Stuhlmann, H., Jähner, D., Wagner, H., Werner, E. and Jaenisch, R. (1983) Retrovirus genomes methylated by mammalian but not bacterial methylase are non-infectious. Nature 304: 275-277.

Singer-Sam, J., Keith, D.H., Tani, K., Simmer, R.L., Shively, L., Lindsay, S., Yoshida, A. and Riggs, A.D. (1984) Sequence of the

promoter region of the gene for human X-linked 3-phosphoglycerate kinase. Gene 32: 409-417.

Soares, M.B., Schon, E., Henderson, A., Karathanasis, S.K., Cate, R., Zeitlin, S., Chirgwin, J. and Efstratiadis, A. (1985) RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. Mol. Cell. Biol. 5: 2090-2103.

Southern, P.J. and Berg, P. (1982) Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. App. Gen. 1: 327-341.

Smiley, J.R., Steege, D.A., Juricek, D.K., Summers, W.P. and Ruddle, F.H. (1978) A herpes simplex virus 1 integration site in the mouse genome defined by somatic cell genetic analysis. Cell 15: 455-468.

Spofford, J.B. (1975) Position effect variegation in Drosophila. Genetics and Biology of Drosophila. M. Ashburner and E. Novitsky (ed), New York, Academic.

Stein, R., Razin, A. and Cedar, H. (1982) In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. Proc. Natl. Acad. Sci. USA 79: 3418-3422.

Stellwagen, R.H. and Cole, R.D. (1969) Chromosomal proteins. Ann. Rev. Biochem. 38: 951-990.

Stewart, C.L., Stuhlmann, H., Jähner, D. and Naenisch, R. (1982) De novo methylation, expression and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA 79: 4098-4102.

Sturm, K.S. and Taylor, J.H. (1981) Distribution of 5-methylcytasine in the DNA of somatic and germline cells from bovine tissues. Nucleic Acids Res. 9: 4537-4546.

Summitt, R.L., Tipton, R.E., Wilroy, R.S. Jr., Martens, P.R. and Phelan, J.P. (1978) X-autosome translocation: a review. Birth Defects: Orig. Art. Ser. 14(6C): 219-247.

Szabo, P., Grzeschik, K.-H. and Siniscalco, M. (1984) A human autosomal phosphoglycerate kinase locus maps near the HLA cluster. Proc. Natl. Acad. Sci. USA 81: 3167-3169.

Takagi, N. (1974) Differentiation of X chromosomes in early female mouse embryos. Exp. Cell Res. 86: 127-135.

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.

Takagi, N. (1978) Preferential inactivation of the paternally derived X-chromosome in mice. 341-360.

Takagi, N. (1980) Primary and secondary nonrandom X chromosome inactivation in early female mouse embryos carrying Searle's translocation T(X;16)16H. Chromosome 81: 439-459.

Takagi, N. and Martin, G.R. (1984) Studies on the temporal relationship between the cytogenetic and biochemical manifestations of X-chromosome inactivation during the differentiations of LT-1 teratocarcinoma stem cells. Dev. Biol. 103: 425-433.

Tani, K., Singer-Sam, J., Munns, M. and Yoshida, A. (1985) Molecular cloning and structure of an autosomal processed gene for human phosphoglycerate kinase. Gene 35: 11-18.

Therman, E. and Patau, K. (1974) Abnormal X chromosomes in man: origin, behavior and effects. Humangenetik 25: 1-16.

Traut, T.W. (1988) Do exons code for structural or functional units in proteins? Proc. Natl. Acad. Sci. USA 85: 2944-2948.

Ullu, E. and Weiner, A.M. (1984) Human genes and pseudogenes for the 7SL RNA component of signal recognition particle. Eur. Mol. Biol. Organ. J. 3: 3303-3310.

VandeBerg, J.L. (1985) The phosphoglycerate kinase isozyme system in mammals: Biochemical genetic developmental and evolutionary aspects. Isozymes: Curr. Top Biol. Med. Res. 12: 133-187.

VandeBerg, J.L., Cooper, D.W. and Close, P.J. (1973) Mammalian testis phosphoglycerate kinase. Nature New Biol. 243: 48-50.

Vanin, E.F. (1985) Processed pseudogenes: characteristics and evolution. Ann. Rev. Genet. 19: 253-272.

Van der Ploeg, L.H.T. and Flavell, R.A. (1980) DNA methylation in the human gamma delta beta-globin locus in erythroid and non-erythroid tissues. Cell 19: 947-958.

Venolia, L., Gartler, S.M., Wassman, E.R., Yen, P., Mohandas, T. and Shapiro, L.J. (1982) Transformation with DNA from 5-azacytidine-reactivated X chromosomes. Proc. Natl. Acad. Sci. USA 79: 2352-2354.

Venolia, L. and Gartler, S.M. (1983) Comparison of transformation efficiency of human active and inactive X-chromosome DNA. Nature 302: 82-83.

Venolia, L., Cooper, D.W., O'Brien, D.A., Millette, C.F. and Gartler, S.M. (1984) Transformation of the HPRT gene with DNA from spermatogenic cells. Chromosoma 90: 185-189.

Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Ann. Rev. Biochem. 55: 631-661.

Weintraub, H., Larsen, A. and Groudine, M. (1981) α-globin gene switching during the development of chicken embryos: expression and chromosome structure. Cell 24: 333-344.

Wigler, M., Levi, D. and Perucho, M. (1981) The somatic replication of DNA methylation. Cell 24: 33-40.

Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedmann, T. and Migeon, B.R. (1982) Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proc. Natl. Acad. Sci. USA 81: 2806-2810.

Wolf, S.F. and Migeon, B.R. (1982) Studies of X chromosome DNA methylation in normal human cells. Nature 295: 667-671.

Wilcox, F.H. and Taylor, B.A. (1981) Genetics of the Akp-2 locus for alkaline phosphatase of liver, kidney, bone and placenta in the mouse. J. Hered. 27: 387-390.

Yen, P.H., Patel, P., Chinault, A.C., Mohandas, T. and Shapiro, L.J. (1984) Differential methylation of hypoxanthine phosphoribosyltransferase genes on active and inactive human X chromosomes. Proc. Natl. Acad. Sci. USA 81: 1759-1763.

Yoshitake, S., Schach, B.G., Foster, D.C., Davie, E.W. and Kurachi, K. (1985) Nucleotide sequence of the gene for human factor IX (antihemophilic factor B). Biochemistry 24: 3736-3750.