THE EFFECTIVENESS OF A PSEUDOGENE PROMOTER IN THE INITIATION OF TRANSCRIPTION.

JONATHAN WHITAKER

Thesis presented for the degree of Doctor of Philosophy

University of Edinburgh

1989

EX BIBL. UNIV. EDINBURGH

To my parents.

ACKNOWLEDGEMENTS

Special thanks to:

Jackie Bogie, Anne Johnson and Jean Gardiner for their help and cooperation in the typing of this thesis.

Frank Johnston for the preparation of photographs.

Raya Al Shawi for micro-injecting, dissections and advice.

Melville Richardson and Ann Duncan for their assistance.

Sam Martin for proof reading the manuscript.

John Bishop for his supervision.

ABSTRACT.

The major urinary proteins (MUPs) of inbred mouse strains are coded for by a family of about 35 genes which are clustered on chromosome 4 (Bishop *et al.*, 1982). Four groups of *Mup* genes which exhibit tissue specific expression and hormonal dependency have been identified (Knopf *et al.*, 1984; Kuhn *et al.*, 1984). The complete sequence of a group 1 gene (BS6) and a group 2 gene (BS2) are known together with 700 base pairs (bp) of 5' flanking sequence (Clark *et al.*, 1985b). The sequence of the group 1 gene is consistent with it being active. The group 2 gene has several lesions and cannot be translated to give a protein with the molecular weight of MUP (Ghazal *et al.*, 1985). This makes it a pseudogene in the context of a group 1 gene. Consistent with this is the observation that no BS2 transcripts are detected in mouse liver RNA. It does not necessarily follow from these observations that the promoter sequences of the group 2 BS2 gene is defective in its capacity to initiate transcription from the BS2 CAP site. This was investigated by implementing standard reverse genetic techniques. Constructs were made containing the SV40 enhancer and early promoter region and approximately 2.2 kilobases (kb) of 5' flanking sequence from BS6 or BS2 linked to the Herpes simplex virus thymidine kinase type 1 gene (HSV*tk*). After transfection of cultured cells with the recombinant plasmids, two assays were used to measure gene expression. Transient levels of *tk* specific mRNA and thymidine kinase enzyme (TK) activity were measured as well as the rate of transformation from a TK^- to a TK^+ phenotype in selective growth medium. Transient

expression was only detectable with constructs containing the BS6 promoter. The transformation efficiency of BS6 constructs was twice that of the BS2 constructs. Individual clones stably transfected with BS6 and BS2 constructs were isolated. Among the clones that carried the BS2 construct the size of the tk mRNA was very heterogeneous. Primer extension and S1 nuclease mapping showed that these transcripts were not initiated at the BS2 CAP site. The clones showed a variety of rearrangements which were consistent with the activation of the tk coding sequences by different cellular promoters. The BS2 promoter had been deleted in all cases. It is concluded that the promoter region of the BS2 pseudogene is not capable of directing accurate transcription in BHKtk$^-$ fibroblast cells. To analyse the effects of the Mup coding regions in a more appropriate environment, the sequences of Mup and thymidine kinase were microinjected into mouse embryos to create lines of transgenic animals. One line expressed HSV thymidine kinase in the liver. The expression was not sexually dimorphic. The size of tk specific transcripts were consistent with the initiation of transcription from the BS2 CAP site. In addition, this line and other G_o males expressed the reporter sequences at high levels in the testis. This tissue does not exhibit any endogenous Mup expression. It is suggested that 2.2 kb of a group 2 Mup gene promoter region is able to direct the synthesis of an HSVtk reporter gene to the liver and testis when removed away from an endogenous silencing element.

ABSTRACT REFERENCES.

Bishop, J.O., A.J. Clark, P.M. Clissold, and S. Hainey. 1982. Two main groups of mouse urinary protein genes, both located on chromosome 4. EMBO. J. 1:615-620.

Clark, A.J., A. Chave Cox, X. Ma, and J.O. Bishop. 1985a. Sequence structure of a mouse major urinary protein gene and pseudogene compared. EMBO. J. 4:3159-3165.

Ghazal, P., A.J. Clark, and J.O. Bishop. 1985. Evolutionary amplification of a pseudogene. Proc. Natl. Acad. Sci. USA. 82:4182- 4185.

Knopf, J.L., J.F. Gallagher, and W.A. Held. 1983. Differential multihormonal regulation of the mouse major urinary protein gene family in the liver. Mol. Cell. Biol. 3:2232-2240.

Kuhn, N.J., M. Woodworth-Gutai, K.W. Gross, and W.A. Held. 1984. Subfamilies of the mouse major urinary protein (MUP) multigene family: sequence analysis of cDNA clones and differential regulation in the liver. Nucl. Acid. Res. 12:6073-6090.

INDEX.

FIGURES.

TABLES.

CHAPTER 1: THE MOUSE URINARY PROTEIN GENES.

Introduction.

The major urinary proteins (MUPs) of inbred strains of mice are coded for by a family of about 35 genes, located on chromosome 4. They are defined as sequences that hybridise with *Mup* specific probes (Hastie *et al.*, 1979; Bishop *et al.*, 1982; Clark *et al.*, 1982). The proteins are acidic and have a molecular weight of about 19500 Daltons. They are synthesised in the liver, submaxillary, lachrymal, sublingual, parotid and mammary glands. Recently expression has been observed in the sebaceous glands of the skin (Held *et al.*, 1989). The rat gene homologue to *Mup*, α2u-globulin, has a similar distribution of tissue specific expression except that it is expressed at high levels in the preputial gland, whereas *Mup* is not (MacInnes *et al.*, 1986). In addition the rat genes are only expressed in the male liver whereas *Mup* genes are expressed in both male and female liver. In each tissue MUP is synthesised as a precursor, secreted into the blood and excreted in the urine. No biological function has been assigned to the *Mup* gene product. However they clearly belong to a protein superfamily, other members of which bind small hydrophobic substances (Sawyer, 1987). These include β-lactoglobulin which binds retinol in several species and human urinary protein (HC) which may mediate the excretion of retinol derived metabolites (Unterman *et al.*, 1981; Pervaiz & Brew, 1985; Ali & Clark, 1988). It has been suggested that MUPs may be involved in binding pheromones thereby facilitating

secretion and excretion of these molecules. Held *et al.*, (1989) show that a *Mup*/SV40 T-antigen transgene was expressed in the sebaceous glands of transgenic mice. This is consistent with the finding of *Mup* mRNA in the skin of transgenic animals. This lends support to the theory that MUPs may have a role in the transport and excretion of small molecules. Although a definite function for MUPs is not established, this multigene family provides an interesting model of developmental and tissue specific multihormonal control of gene expression.

Mup gene expression is detected in the liver of mice between 2 and 4 weeks of age (Barth *et al.*, 1982). This time period corresponds to the onset of sexual maturity and an increase in the circulation of testosterone. Consistent with this observation, Held *et al.*, (1989) found expression of a *Mup*/SV40 T-antigen transgene in the liver of mice at around 3 weeks.

Mup genes can be divided into four groups.

Analysis of the structure of the specified protein and comparison of restriction sites and nucleotide sequence show that there are four main phylogenetic groups of *Mup* genes (Al Shawi *et al.*, 1989, in press):

Group 1 genes.

Group 1 genes are expressed in the liver and are detectable at low

levels in the prelactational mammary gland (Clissold & Bishop, 1982; Shahan et al., 1987a). The liver synthesises a great deal of Mup mRNA. This constitutes about 5-10% of total mRNA in the adult male liver and about 1-2% in the adult female liver (Derman, 1981). It has been calculated that approximately 90% of the Mup mRNA in the male mouse liver corresponds to group 1 transcripts (Clark et al., 1984a; Kuhn et al., 1984). To date only four group 1 genes are known to be expressed in BALB/c mice; cLiv1 and cLiv6 (Shahan et al., 1987a); p1057 (Kuhn et al., 1984) and pMUP(11) (Clark et al., 1985a). Clark et al., (1982) describe the sequences of four group 1 genes isolated as genomic clones from BALB/c DNA. They are named BS6, BS5, BS1 and BL1. These sequences show a nucleotide homology of approximately 98%. Sequence analysis has shown that cLiv1 is identical to BS1 and cLiv6 is identical to BL1. The pMUP(11) clone appears to be the same as the BS6 genomic clone (J. Bishop, personal communication).

Shahan et al., (1987a) isolated the cDNA clones cLiv1 and cLiv6 from male and female liver libraries respectively. Four out of five clones isolated from the male liver library were identical to cLiv1 (BS1) and eight out of nine from the female library were identical in sequence to cLiv6 (BL1). This work proposes that these transcripts represent the major Mup mRNA species in the liver of male and female mice. There is evidence that each of these genes is present in the BALB/c genome in a single copy (Al Shawi et al., 1989, in press). Hybridisation with oligonucleotide probes was used to investigate the RNA extracted from the livers of both sexes

(Shahan et al., 1987b). These probes did not discriminate between cLiv1 and cLiv6. The results are inconclusive in determining whether each gene is expressed exclusively in one sex or the other.

Using gene specific probes McIntosh & Bishop (1989) show that the group 1 genes BS1 and BL1 are expressed at different levels in male and female mice. Probing of mRNA transcripts with group 1 gene specific probes has shown that BL1 is the most abundantly expressed Mup gene in male and female livers. This is in contradiction to the work of Shahan and colleagues (1987). The BL1 RNA constitutes approximately 60% of male and 75% of female total Mup RNA. BS1 RNA was also detected. In the male this occurred at levels which were one third that of BL1. In the female, BS1 transcripts are detectable but at a much lower level compared to BL1. The levels of both RNAs in the female liver are raised to male-like levels when testosterone is administered continuously by a subcutaneous pellet. Neither RNA was detected in the prelactational mammary gland.

The group 1 gene BS6 (pMUP(11)) is also expressed at different levels in male and female mice. Al Shawi et al., (1988) analysed the expression of a Mup/tk transgene containing 2.2 kb of BS6 5' flanking region linked to the coding region of a Herpes simplex virus type 1 thymidine kinase gene (HSVtk). In three out of five lines the levels of HSVTK in the liver had a similar male to female ratio of 10.

Group 1 genes are expressed in the mammary gland.

Shahan et al., (1987b) isolated 5 cDNA clones from a prelactational mammary gland cDNA library. Two of these clones were sequenced and found to be identical to the cLiv6 clones isolated from the female liver library. It has been proposed that the BL1 gene is the most abundantly expressed group 1 gene in the mammary gland. In contrast McIntosh & Bishop (1989) failed to detect any BL1- or BS1-specific transcripts when RNA from this tissue was hybridised with the two gene-specific probes.

Group 2 genes.

Group 2 genes are pseudogenes in the context of group 1 genes. They include the genes BS2, BS102, BS109 and BL25 (Clark et al., 1982). They are distinguishable from the group 1 genes on the basis of nucleic acid hybridisation experiments (Bishop et al., 1982). Sequence comparisons of the first exon of four group 2 genes show a nucleotide homology of 98% and the presence of a nonsense mutation (TGA) at amino acid 7 (Ghazal et al., 1985). The sequence cannot be translated to give a protein with the molecular weight of MUP. Other lesions are also present. BL25 contains a stop codon in place of amino acid 2 and BS2 contains a second stop codon and a frameshift mutation (Clark et al., 1985b). It has been suggested that although the nonsense mutation does not allow the genes to encode a full length MUP protein, a smaller peptide could be a

viable translation product (Clark et al., 1984a).

Shahan et al., (1987b) constructed an oligonucleotide probe derived from a sequence conserved in the four group 2 genes and which spans the nonsense mutation. No complementary RNA was detected when this was hybridised to RNA from several Mup expressing tissues. This result was confirmed with a group 2 gene specific oligonucleotide probe on different RNA samples (McIntosh & Bishop, 1989). Thus it appears that the group 2 genes are not transcribed into stable RNA. The sequence of a group 1 gene (BS6) and a group 2 gene (BS2) have been published (Clark et al., 1985b). They show a 13% divergence over the full length of the gene.

Group 3 genes.

These genes are also expressed in the liver. The first to be described and assigned to this group was MUP15 isolated from a liver cDNA library (Clark et al., 1985a). The MUP15 gene has diverged from the group 1 consensus sequence by approximately 15%. A second gene, MUP16, isolated from a genomic clone shows around 5% divergence from MUP15. MUP15 also differs from the group 1 genes in having a longer signal peptide sequence and a different splice configuration between exons 6 and 7. Unlike the group 1 sequences, MUP15 contains a N-linked glycosylation site and has been shown to code for a glycosylated protein detectable in mouse urine (Kuhn et al., 1984; Clark et al., 1985a). MUP16 also encodes a protein with the same glycosylation site. It differs from all Mup genes

cloned to date because of the lack of exon 7 and a large 3' deletion (Ma, 1987). Shahan et al., (1987a) isolated sequences identical to MUP15 in cDNA clones cLiv4 and cLiv7 from male and female liver cDNA libraries respectively. On the basis of the frequency of isolation of cDNA clones, it is estimated that this species constitutes about 10-20% of Mup mRNA in the liver of both sexes.

Group 4 genes.

This group constitutes Mup genes that are expressed predominantly in the lachrymal and salivary glands. Lachrymal gland Mup mRNA constitutes 0.5-1% of total lachrymal gland mRNA synthesised in adult males (Shahan & Derman, 1984). In vitro translation of this mRNA gives a protein which is significantly more basic than other MUPs (Shaw et al., 1983). Shahan et al., (1987a) constructed cDNA clones from Mup-specific mRNA isolated from the lachrymal gland. Each of the 11 cDNA clones isolated was identical to a clone cLac1 indicating that all of lachrymal Mup mRNAs are transcribed from this one gene. The unique sequence of cLac1 confirms that a distinct Mup gene is expressed in the lachrymal gland. Clark et al., (1982) isolated a gene BL2 from a partial genomic clone. Extensive restriction enzyme site homology between BL2 and cLac1 suggest that they are the same gene.

In the submaxillary gland, Mup mRNA constitutes 0.2-0.5% of the total mRNA (Shahan & Derman, 1984). Mup cDNA clones were constructed from the submaxillary RNA of prepubescent male mice.

Each was identical in sequence to a single gene isolated in a cDNA clone cSmx1 (Shahan et al., 1987a) showing that only one Mup gene is actively expressed in this tissue.

Specific oligonucleotide probes were constructed from sequences of variable segments in the cloned genes. Probing of mRNA from various tissues confirms that cLac1 and cSmx1 are not expressed in the liver (Shahan et al., 1987a). However, mRNA complementary to the cLac1 probe was detected in the lachrymal gland and in low levels in the parotid gland. Transcripts with affinity for the cSmx1 probe were detected in the submaxillary gland and in low levels in the sublingual gland.

The group 4 cDNA clones differ in nucleotide sequence from group 1 genes by approximately 10%, from group 2 genes by 12% and from each other by 7%.

Summary.

Tissue-specific synthesis of Mup mRNA is thus brought about by two mechanisms: the expression in different tissues of different members of the family and the expression of a single gene at various levels in different tissues. A summary of the expression of different Mup genes in different tissues is shown below. The names of clones are referenced to Clark et al., (1982) and (Shahan et al., 1987a).

Summary.

TISSUE	*Mup* GENE GROUP			
	1	2	3	4
Liver	BS1 (cLiv1)		MUP15 (cLiv4)	
	BL1 (cLiv6)		(cLiv7)	
	pMUP(11)/BS6		MUP16	
	p1057			
Mammary	BL1 (cLiv6)			
Submaxillary				(cSmx)
Lachrymal				BL2/(cLac1)
Sublingual				(cSmx)
Parotid				BL2/(cLac1)

Mouse urinary protein gene organisation.

Each *Mup* gene is about 3.9 kb long. Most of the genes fall into two main groups, 1 and 2, on the basis of nucleic acid hybridisation experiments (Bishop *et al.*, 1982). Between 13 and 15 group 1 genes and the same number of group 2 genes are arranged in a pairwise divergent orientation. Each gene is associated with 5 kb of homologous 5' flanking sequence and 11 kb of 3' flanking sequence that contains regions of homology interspersed with non homologous sequences. The two 5' regions of each pair are separated by 6 kb of

non-duplicated DNA. This organisation forms a set of giant imperfect palindromes each about 45 kb in size (Clark et al., 1984b; Bishop et al., 1985; see Figure 1A). Sequence relationships between different palindromes show that they have evolved from a common ancestor. This implies that the 45 kb palindrome is not only the predominant unit of Mup gene organisation but also the unit of Mup gene evolution (Ghazal et al., 1985). The remaining 5 to 9 Mup genes do not seem to be incorporated into palindromic structures. Of these at least 3 are pseudogenes that have resulted from a number of different rearrangements (Clark et al., 1984b). Thus the number of active Mup genes in the mouse genome is about 15 (Bishop et al., 1982; Shahan et al., 1987b), roughly the same as the total number of different proteins synthesised in all Mup expressing tissues.

Mup gene structure.

The 3.9 kb transcription unit of the Mup genes contains 7 exons (Clark et al., 1984a). The first 6 exons contain the coding sequences while the last exon consists entirely of non coding sequences. Three different splicing configurations have been found which result from the presence of alternative splice sites within the untranslated region of exon 6 (Clark et al., 1984a, 1985a; see Figure 1B). The most abundant liver Mup transcripts contain part of exon 6 and all of exon 7. The smaller and less abundant (by a factor of about 10) liver Mup transcripts contain an extended exon 6 but completely lack exon 7. The main processed transcription

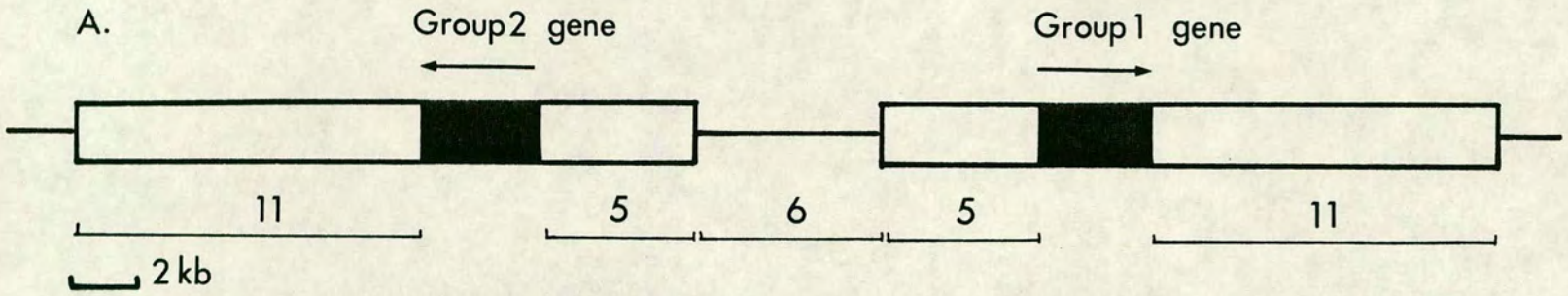
Figure 1.

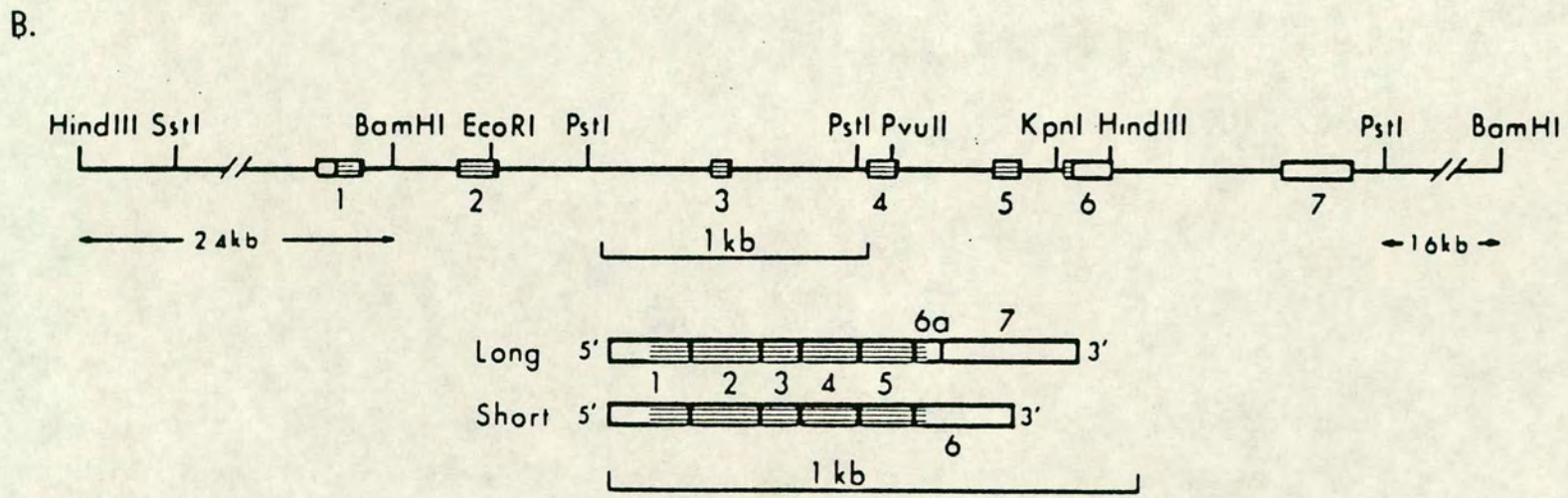
A) Organisation and structure of the Mup genes.

Diagramatic representation of the 45 kilo base unit. Group 1 and group 2 genes are approximately 3.9 kb long (shaded). Open rectangles are regions of homology between the flanking regions of the two groups of genes. The homology is not continuous over the 11 kb 3' flanking region but is interrupted by regions of non-homology. Genes are in a head to head (divergent) orientation. The 45 kb gene pair structure constitutes the evolutionary unit of the Mup locus (see text).

B) A diagram to show the two main forms of the Mup mRNA.

The two main Mup RNAs showing their structural relationship to the gene BSu6 are shown. Short mRNA contains exons 1-6. Long mRNA contains exons 1-5, a short 5' region of exon 6 and the whole of exon 7. The greater part of liver Mup mRNA is in the long form and nearly all of this is transcribed from the group 1 genes. Exons are shown as boxes and the coding regions are shaded. Adopted from Clark et al., 1984a.


A.
Group2 gene
Group1 gene
11
5
6
5
11
2kb


B.
HindIII SstI
BamHI EcoRI
PstI
PstI PvuII
KpnI HindIII
PstI
BamHI
1
2
3
4
5
6
7
2.4kb
1kb
1.6kb
6a
7
Long
5'
3'
1
2
3
4
5
Short
5'
3'
6
1kb

products of the Mup genes, long RNA and the two short forms, are 878, 759 and 741 nucleotides long respectively. The long RNA is the prevalent product of slicing of the group 1 gene transcripts (Clark et al., 1984a). In contrast the transcript from pMUP15 is spliced to give an extended exon 6 and subsequently a shorter RNA (Figure 1B). The mRNA specifying sequences of 4 group 1 genes and extensive nucleotide sequences of 4 group 1 cDNAs are known (Clark et al., 1985a, 1985b). These code for a signal peptide 18 amino acids long and a mature protein 162 amino acids long. In contrast, the signal peptides of four group two genes are all different. They vary in length from 19 to 25 amino acids.

The evolution of Mup genes.

The predominant organisation of the Mup locus is an array of 45 kb domains consisting of group 1 and group 2 genes (Figure 1A). Ghazal et al., (1985) describes the nucleotide sequence of the first exon of six group 1 and four group 2 genes and derives a consensus sequence for each. The consensus sequences differ by 17%. One of the differences is between a glycine (GGA) in the group 1 sequence and a stop codon (TGA) in the group 2 sequences at amino acid 7. This lesion, common to all group 2 genes, is their most significant feature. It implies that the mutation was present in a pseudogene which was ancestral to all four group 2 genes.

The complete nucleotide sequence of genes BS6 and BS2 have been established (Clark et al., 1985b). The divergence between the two

genes is high, approximately 10%. However, within a group the genes are very homogenous and differ from each other by only 1-2%. To explain this phenomenon an evolutionary model based on the 45 kb unit has been proposed (Ghazal et al., 1985). Selection acting against an unfavourable newly arisen, actively transcribed group 1 gene would lead to the elimination of the 45 kb unit which contained it. A similar action would apply to the group 2 genes until the pseudogene lesion occurred, then further mutations could accumulate more rapidly. The two main parts of the amplification unit cannot replace another unit. This leads to homogeneity of the units and uniformity of group 1 and group 2 genes. To account for the tolerance of the group 2 gene it has been proposed that the generation of the gene pair unit arose at about the same time as the pseudogene mutation.

CHAPTER 2: REGULATION OF MUP GENE EXPRESSION.

Sexually dimorphic expression of Mup genes.

In BALB/c male mice, *Mup* mRNA is the most abundant class of liver mRNA whilst in the female the most abundant species is serum albumin (Clissold & Bishop, 1981). In BALB/c adult male mice, *Mup* mRNA constitutes about 8% of the total liver polyadenylated mRNA. This level is five times higher than that of adult female mice (Hastie & Held, 1978; Hastie *et al.*, 1979), although expression in females may be rather variable (McIntosh & Bishop, 1989). The different levels of *Mup* mRNA in male and female liver reflects different rates of transcription as determined by run-on transcription experiments (Derman, 1981). Sexually dimorphic expression of a group 1 BS6/HSV*tk* transgene is observed in transgenic mice (Al Shawi *et al.*, 1988). The male to female ratio was about 10 in 3 out of 5 transgenic lines. Run-on experiments showed that the different levels of HSV*tk* RNA and thymidine kinase observed are to some extent mediated by different rates of transcription. Held *et al.*, (1989) also observed differences between the sexes in the expression of a *Mup*(BL6)/SV40 T-antigen transgene although no run-on experiments were conducted in this case.

The lachrymal gland also exhibits sexual dimorphism with respect to *Mup* gene expression. The male lachrymal gland has 5 times as much *Mup* mRNA as female (Shaw *et al.*, 1983). The submaxillary gland does not show sexual dimorphism with respect to *Mup* expression.

The effect of testosterone on Mup gene expression.

Early studies on the regulation of Mup genes came from experiments which revealed a group of urinary proteins that exhibit both sex- and strain- specific phenotypes. The administration of testosterone to females resulted in an increased amount of excreted MUPs and a shift to the male phenotype of expressed proteins (Finlayson et al., 1965). Livers from female and castrated male mice have reduced levels of Mup mRNA, up to one fifth of normal males. This can be elevated to normal male levels upon administration of testosterone (Clissold et al., 1984). In female mice transgenic for a BS6/tk transgene (Al Shawi et al., 1988) the presence of testosterone was sufficient to induce increased levels of thymidine kinase, but not to male levels. The data suggests that the increase in transcription is mediated by the action of testosterone, via its receptor, on sequences within the 2.2 kb BS6 promoter region.

It has been proposed (Norstedt & Palmiter, 1984) that testosterone acts on the pituitary to induce a discontinuous pattern of growth hormone production in males, compared with a continuous production in females. This determines the sexual dimorphism of the liver. It is also suggested that a discontinuous stimulation of growth hormone receptors causes a discontinuous release of insulin-like growth factor I (IGF-I). This release induces MUP synthesis. A continuous stimulation of growth hormone receptors (as in the female) up-regulates growth hormone/prolactin receptors resulting in a

continuous release of IGF-I. This results in the down-regulation of IGF-I receptors and a low production of MUP.

The action of other hormones on Mup gene expression.

Regulation of Mup gene expression has been studied in C57BL/6J mice by the administration of hormones to mutants or surgically altered mice with hormone deficiencies (Knopf et al., 1983; Kuhn et al., 1984). To examine the effect of growth hormone, the mutant mouse strain little was used. The pituitary glands of these mice are insensitive to growth hormone releasing factor and very low levels of growth hormone are detected in the serum (Jansson et al., 1986). Mup mRNA levels are reduced over 100-fold in these mice but are restored to male levels on the administration of growth hormone (Knopf et al., 1983). The action of growth hormone in the absence of other pituitary controlled hormones was studied in hypophysectomised mice which lack pituitary function. Synergistic effects of hormones are indicated by the absence of testosterone-mediated Mup mRNA synthesis and the non additive effect of administration of both growth hormone and thyroid hormone. Testicular feminised tfm/Y mice, which lack a functional major testosterone receptor protein and are relatively androgen insensitive, were used to further test the testosterone independent induction of Mup mRNA by growth hormone and thyroxine. Cooperativity of the two hormones in producing an elevated level of Mup mRNA was again observed. These results together indicate that Mup genes expressed in the liver are under complex multihormonal

control. Further evidence comes from experiments using thyroidectomised female mice since the administration of each of the hormones testosterone, growth hormone and thyroxine results in increased levels of *Mup* mRNA in the liver and different patterns of mRNA translation products.

Previous reports indicate that rat α2u-globulin expression in the liver is regulated by glucocorticoids and that cloned genes introduced into mouse L cells are regulated by dexamethasone (Kurtz 1981). Held *et al.*, (1987) transfected L cells with various genomic clones of *Mup* isolated from C57BL/6J and BALB/c mice. A low induction of MUP synthesis by glucocorticoid was observed but this may have been related to cell physiology. The regulation by these hormones contribute to the sexual dimorphism of *Mup* expression in the liver (Szoka & Paigen, 1978; Knopf *et al.*, 1983; Norstedt & Palmiter, 1984). Mice transgenic for an α2u-globulin genomic clone containing 2 kb of upstream sequence express α2u-globulin in adult male liver and preputial gland. Expression can be induced by glucocorticoid treatment of ovariectomised mice (da Costa Soares *et al.*, 1987). The upstream sequences present in the transgene include the elements identified by Addison & Kurtz (1986) as being involved in the glucocorticoid induction of α2u-globulin gene expression in L cells. This suggests that these elements may also be involved in the glucocorticoid regulation of α2u-globulin in the rat liver, first demonstrated by Kurtz & Feigelson (1978).

In the lachrymal gland, testosterone induction tends to be

independent of growth hormone or thyroxine. Mup gene expression in the submaxillary gland does not appear to be under hormonal control. The hormonal regulation of mammary gland Mup genes has not been studied. It is not known whether the hormonal modulation of MUPs is by direct or indirect action in the particular tissues.

Tissue-specific expression of MUPs.

Transgenic mice have been used to analyse DNA sequences required for Mup tissue specific expression. Al Shawi et al., (1988) and Held et al., (1989) show that the gene sequences necessary for tissue and temporal control of Mup genes BS6 and BL6 are located within the 5' flanking regions of these genes. Al Shawi et al., directed the expression of a Herpes simplex virus thymidine kinase (HSVtk) reporter by 2.2 kb of 5' BS6 (group 1) sequences. Expression was observed in the liver of male and female mice. In addition the regulation of expression by testosterone exhibited sexual dimorphism. These are the characteristics of a group 1 Mup gene. Held et al., (1989) analysed the tissue specific expression of a group 1 Mup (BL6)/SV40 T-antigen hybrid gene. The 5' region constitutes 2.5 kb. Expression was observed in male and female livers. Preliminary results (Al Shawi et al., unpublished) suggest that liver expression is directed by sequences located more than 360 bp upstream from the Mup CAP site. HSVtk expression in the liver was abolished in mice transgenic for this short DNA fragment (see Chapter 11).

CHAPTER 3: THE EUKARYOTIC PROMOTER.

Introduction.

A general strategy for identifying *cis* regulatory elements has been termed reverse genetics. Cloned genes are truncated or mutated *in vitro* and their expression is then studied after transfection into cell lines. Detailed molecular analysis of a number of different promoters reveals a common pattern of organization in genes transcribed by RNA Polymerase B (Grosschedl & Birnstiel, 1980,1983; Benoist & Chambon, 1981; Dierks *et al.*, 1981,1983; McKnight *et al.*, 1981; Everret *et al.*, 1983; McKnight & Kingsbury, 1983). Definable structural features include the *CAP site* which identifies the sequence corresponding to the 5' terminus of the messenger RNA (Schenk, 1981). At the start point there is no extensive homology of sequence but there is a tendency for the first base of mRNA to be an adenine residue flanked on either side by pyrimidines (Breathnach & Chambon, 1981). This consensus may not be a true reflection of CAP structure as it was collated from only 22 genes. The *TATA* consensus sequence is located at approximately 30 bp upstream from the CAP site. It appears to position the location of the CAP site. Deletion of the TATA box (with retention of the upstream sequences) generates messages with heterogeneous 5' ends as seen in the Sea Urchin Histone H2A gene (Grosschedl & Birnstiel, 1980) the SV40 early genes (Benoist & Chambon, 1981) and the rabbit β-globin gene (Grosveld *et al.*, 1982). In the case of the SV40 early region, the amount of early mRNA produced does not seem to be

strongly influenced while deletion in both the histone and β-globin genes is a strong down mutation. Deletion of the CAP site does not abolish transcription but leads to the generation of new initiation sites with lower transcriptional efficiencies. Exceptions to the universal presence of the TATA box sequence are the Papovavirus late genes and adenovirus-2 E2 gene (Baker & Ziff, 1981) and the so called "housekeeping" genes. A number of other elements which appear to affect the efficiency of transcription have been categorized. These include the CCAAT consensus located at approximately 80 bp upstream from the CAP site and regions up to 120 bp away. These contain imperfect repeats of 12 to 15 bp of GC-rich sequences (Dierks et al., 1981, 1983; Everett et al., 1983). Mutagenesis studies suggest that the strength of the promoter is determined by the number and type of upstream promoter elements which seem to act regardless of orientation. More distant sequences which influence transcription are termed enhancers. Originally identified in SV40 (Moreau et al., 1981; Banerji et al., 1981) they are cis-acting DNA segments which dramatically stimulate transcription from RNA polymerase B class homologous or heterologous promoters. They act in an orientation-independent manner and over long distances. Tissue-specific enhancers have recently been identified often lying a few kb away from the CAP site. Examples include the mouse albumin (Pinkert et al., 1987) and the chicken ovalbumin genes (Kaye et al., 1986). In contrast to the SV40 enhancer which is active in a variety of cell lines, some enhancers exhibit a host cell preference (Spandidos & Wilkie, 1983; De Villiers et al., 1982; Yoshimura et al., 1985) and strict cell lineage specificity (Queen & Baltimore,

1983). Silencers are thought to have a dominant role in tissue-specific gene expression. They have been used to explain the mis-expression of Mup genes in transgenic mice (Al Shawi et al.,1988; see Chapter 11). Specific trans-acting proteins are thought to interact with silencer sequences to block transcription. This is analogous to the way in which prokaryotic repressors interact with operators, for example Gal80 and the galactose operon. It is generally agreed that in combination with specific cis-acting DNA elements which bind specific regulatory proteins, these various regions constitute a promoter for eukaryotic gene transcription.

DNA elements as recognition sites for trans-acting proteins.

The Sp1 binding site.

Proteins that interact with specific cis DNA recognition elements are thought to contribute to the regulation of gene expression. Recently many have been identified by gel retardation and footprinting assays. Mutational analysis of the SV40 early promoter in HeLa cells revealed the existence of a promoter element containing the sequence GGGCGG which binds a trans-acting factor designated Sp1 (Dynan & Tjian, 1983a). DNAase I footprint and protection studies on mature or mutant SV40 early promoters have shown that Sp1 binds in a sequence-specific manner to the GC rich hexanucleotides that occur twice in each 21 bp repeat of the SV40 early promoter (Dynan & Tjian, 1983b; Gidoni et al., 1985; see

Figure 2A). Each binding site occurs once per turn of the DNA helix. When the promoter is fully occupied, 5 Sp1 factors lie predominantly on the same face of the DNA helix. The Sp1 recognition site that is closest to the RNA initiation site is typically located 40-70 nucleotides upstream of the start and despite the asymmetry of Sp1 binding sites, they are functional in either orientation. Kadonaga et al., (1987) showed that purified Sp1 (Briggs et al., 1986) requires zinc(II) for sequence specific binding to the DNA. This is a feature characteristic of many activator proteins (Chapter 4).

Sp1 has been shown to be involved in regulating the expression of a number of different genes. These include the Herpes simplex virus thymidine kinase gene (McKnight & Tjian, 1986), human metallothionein IIA gene (Lee et al., 1987a), mouse hydrofolate reductase gene (Dynan et al., 1986) and the human heat shock hsp70 gene. Sp1 binding sites are often found near binding sites for other transcription factors, for example with the CTF/NF-1 factor in the HSVtk gene (Jones & Tjian, 1985) and Ap-1 (Lee et al., 1987a). This suggests that these factors may act in conjunction with each other to modulate the level of transcription. DNAase footprint experiments indicate that Sp1 binds at two GC hexanucleotide domains within the HSVtk promoter. One centered within a domain ds1 and another, in opposite orientation, in the ds11 domain (see Figure 2B).

CpG islands as gene markers. The role of Sp1.

Vertebrate genes, in contrast to invertebrates, are nearly always methylated to some extent. A small fraction of the vertebrate genome is characteristically non-methylated and constitutes discreet DNA domains, usually 1-2 kb long, containing an unusually high number of CpG dinucleotides relative to the bulk of the genome (Bird, 1986; 1987). Many CpG islands are associated with genes eg housekeeping genes, particularly at the start site of transcription (Bird, 1986) although in a few tissue-specific genes they appear to be present entirely downstream of the CAP site. Therefore CpG islands not only mark genes but are likely to contain their promoters and 5' ends. The effect of methylation on chromatin structure remains to be resolved. However, it has been suggested that the frequent association of these CpG islands at the 5' end of a gene reflects a preferred site of interaction between DNA and DNA binding proteins. These proteins would be prevented from binding to other stretches of DNA because of methylation. One candidate for this model would be Sp1 as its consensus contains a CpG dinucleotide and is GC-rich. However Dynan (1989) reviews work which shows that the methylation of Sp1 sites does not always inhibit Sp1 directed transcription.

The CCAAT motif as a recognition sequence for cellular transcription factors.

A region of approximately 24 bp containing a binding site for a protein factor distinct from Sp1 is positioned between the two Sp1

binding sites of the HSVtk promoter (Figure 2B). This protein has been isolated from (i) HeLa cells (Jones et al., 1985b) and (ii) rat liver nuclei (Graves et al., 1986). Since the pentanucleotide CCAAT occurs within the binding domain, the protein factor has been termed CCAAT transcription factor CTF (i) or CCAAT binding protein CBP (ii). The CCAAT sequence is required for the transcription of a class of cellular genes that include many vertebrate globin gene families. The CCAAT element of the mouse β-globin promoter interacts with yet another promoter element that contains the sequence CCACACCCG whereas the CCAAT element of the human α-globin promoter appears to function independently (Charney et al., 1985; Myers et al., 1986). The CCAAT element is also found in the heat shock elements of Xenopus and human hsp70 genes (Beinz & Pelham, 1986) and in the 5' region of the mouse albumin gene (Lichensteiner et al., 1987). Although various CCAAT containing elements appear similar in sequence, evidence suggests that different proteins are capable of distinguishing between them (Dorn et al., 1987). Chodosh et al., (1988) describes three related but distinct proteins from HeLa cells designated CP1, CP2, and NF-1 which specifically recognize particular subsets of CCAAT elements. Binding sites for CTF in human Ha-ras and the α-globin promoter were highly homologous to sequences which recognised the nuclear factor NF-1. This is required for initiation of adenovirus DNA replication in vivo (Bernstein et al., 1986). Jones et al., (1987) have since shown these factors to be indistinguishable at the polypeptide level.

The TATA box binding factor TF11D.

Moncollin et al., (1986) and Fire et al., (1984) have shown that at least four general transcription factors BTF1, BTF2, BTF3 and STF are absolutely required for accurate initiation by RNA polymerase B from the Adenovirus 2 major late promoter (Ad2MLP). The TATA box factor BTF1, interacts with a large region of the promoter, including the TATA box and initiation start sites, to form a stable pre-initiation complex (Shi et al., 1986; Zheng et al., 1987). This interaction occurs in the absence of RNA polymerase B and is facilitated by a factor STF. Two other factors BTF2 and BTF3 are then required for accurate initiation. Cavallini et al., (1988) showed that a yeast factor, GAL4, can substitute for the HeLa cell BTF1 TATA box factor. This indicates the striking evolutionary conservation of trans-acting regulatory proteins and suggests that BTF1 may be a ubiquitous factor.

Hormone responsive elements.

The hormone responsive elements are closely related 15 bp palindromes. The dyad symmetry of these elements suggests that nuclear receptors may bind as a dimer, with each receptor molecule recognising an arm of the element. Comparison of several functional glucocorticoid responsive elements (GREs) has indicated a consensus, 5'-GGTACANNNTGTTCT-3', which differs in only 2 bp from that of the oestrogen receptor (Green & Chambon, 1988). While no clear consensus sequence has yet been defined for the progesterone and androgen

receptor elements, a GRE 15-mer is able to mediate induction by progesterone, androgens and mineralocorticoids (Beato et al., 1989). This suggests that the hormone responsive elements are a family of related sequences which have co-evolved with a receptor family. In terms of structure and function it seems that steroid hormone receptors can be divided into two groups: the glucocorticoid, mineralocorticoid and androgen receptors being more closely related to each other than the oestrogen, thyroid, retinoic acid and vitamin D receptors (Chang, 1988 and references therein).

The SV40 enhancer and early promoter region.

Enhancers also contain short discrete DNA sequences which specifically interact with proteins. In Simian Virus 40, an AT-rich region which bears similarity to the TATA box of eukaryotic genes, is believed to fix the early mRNA start sites (Ghosh et al., 1981). Flanking this region are three direct repeats of a GC-rich 21 bp sequence. Each contains two GC hexanucleotide sequences which bind the transcription factor Sp1. Complete deletion of these repeats severely affects early promoter function (Everett et al., 1983). Two additional 72 bp repeats are positioned upstream. Removal or disruption of both these elements reduces early transcription, but deletion of a single repeat does not have a substantial effect. Unlike the 21 bp repeats, the 72 bp elements can activate transcription when placed up to 1 kb either side of a transcription unit (Moreau et al., 1981). These properties are consistent with those of an enhancer element. However systematic mutagenesis of the

SV40 enhancer region (Zenke et al., 1986) shows that the 72 bp sequence is not the SV40 enhancer. Enhancer activity is confined to a 100 bp region, which includes the 72 bp sequence and essential upstream sequences. It can be dissected into two domains A and B (Maniatis et al., 1987; see Figure 2A). These domains exhibit little enhancer activity on their own. Irrespective of relative orientation and to some extent the distance between them, their association results in a 400-fold enhancement of transcription. Enhancer activity can also be generated by duplication of either domain. Another study identified 3 separate domains A, B and C (Herr & Clarke, 1986; Clarke & Herr, 1987). The A and C domains together constitute the B domain named by Zenke above (Jones et al., 1988). The activity of each domain is in turn dependent on the integrity of specific sequence motifs (Figure 2A). These include the P element which binds the transcription factor Ap-1 (Lee et al., 1987a). The GT1 and GTII domains are highly homologous but bind specific proteins such as Ap3 (GTI) and Ap5 (GTII) (Jones et al., 1988). The enhancer also contains a direct repeat of the Sph/Octamer motif ATTGCAT which binds a range of nuclear proteins (Jones et al., 1988; Sturm et al., 1987). The host range of enhancer activity may depend on the particular motifs present since the corresponding factors may be tissue specific. Consistent with this theory, Davidson et al., (1986) have shown that different sequence motifs within the SV40 enhancer are recognised by different factors present in B cell and HeLa cell nuclear extracts. Schirm et al., (1987) have shown that all of the sequence motifs may have different properties in different cell lines.

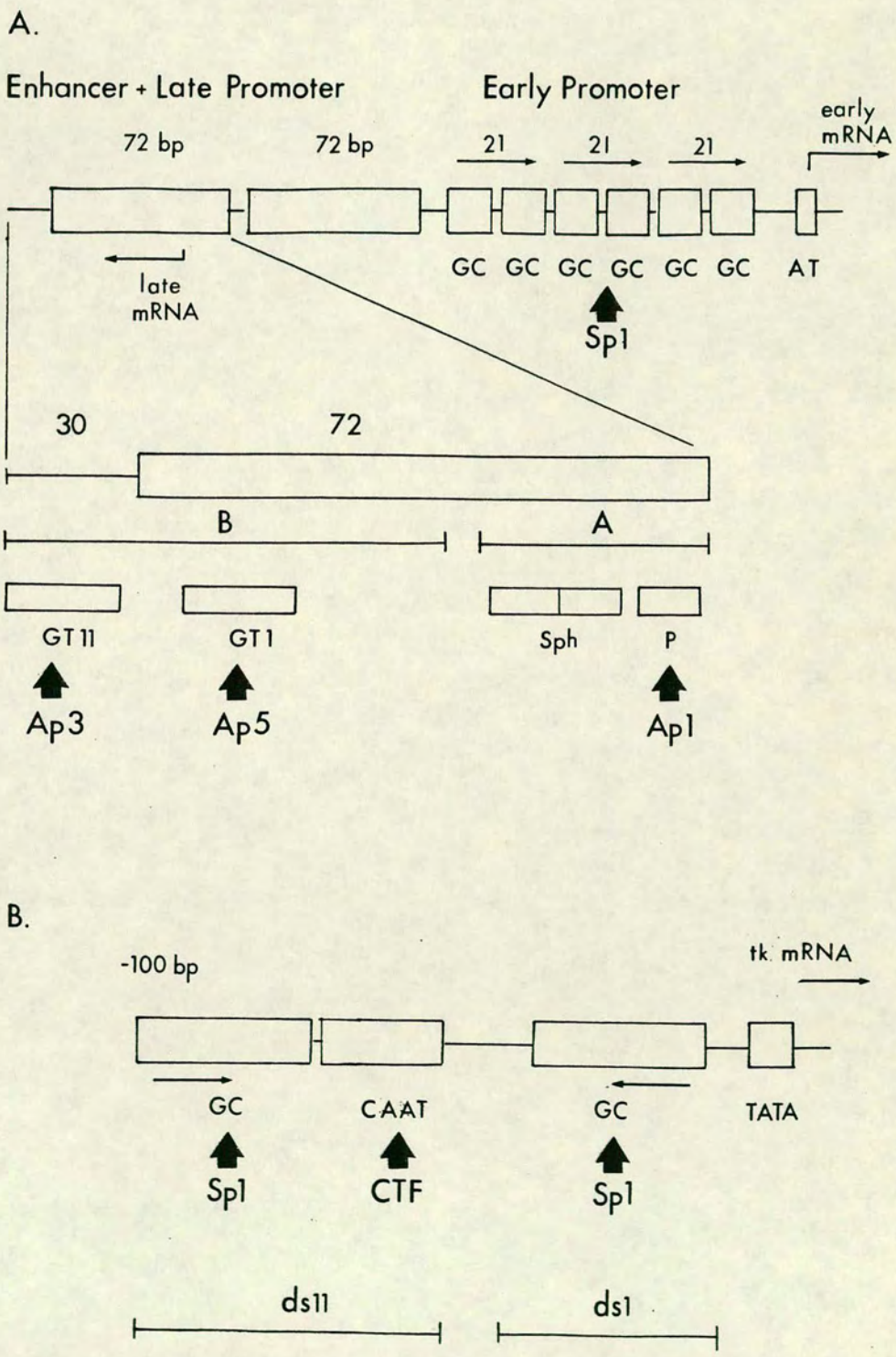
Figure 2.

(A) Cis-acting control sequences of the SV40 promoter.

The SV40 early promoter constitutes three 21 bp repeats each of which contains two GC hexanucleotide motifs which bind the transcription factor Sp1. The 72 bp repeats harbour *cis* regulatory elements for enhancer activity and late promoter funtion. The SV40 enhancer spans 100 bp. It is composed of two domains A and B which contain DNA specific elements recognised by a variety of cellular proteins (Zenke *et al.*, 1986).

(B) A schematic diagram of the cis acting control sequences of the HSVtk gene.

Cis-acting control sequences of the HSV*tk* gene include a TATA box and a cellular transcription factor (CTF) binding site sandwiched between two GC hexanucleotide Sp1 binding sites. The sites fall into two functional domains ds1 and ds11 as shown. (Adopted from McKnight & Tjian, 1986).


A.
Enhancer + Late Promoter
Early Promoter
early mRNA
72 bp
72 bp
21
21
21
late mRNA
GC GC GC GC GC GC AT
Sp1
30
72
B
A
GT 11
GT 1
Sph
P
Ap3
Ap5
Ap1
B.
-100 bp
tk. mRNA
GC
CAAT
GC
TATA
Sp1
CTF
Sp1
ds11
ds1

Summary.

The activity of promoters and enhancers is determined (i) by the types of control elements present and (ii) by the interaction of the regulatory proteins that recognise these control elements. A considerable degree of promoter diversity can be generated by combination of a relatively small number of trans-acting elements on cis-acting sequences. How the factors that recognise enhancers can act on promoters over a large distance has been proposed in several models (Serfling et al., 1985; Schaffner, 1989). Recent observations are consistent with a model in which transcription is stimulated by interactions between proteins bound to the promoter and enhancer with the looping out of the intervening DNA (Ptashne et al., 1986). The binding of other repressors and activators, DNA polymerase is then mediated by this initial looping out. A second hypothesis for transcriptional activation is that the binding of activators causes some local change in chromatin structure which renders its TATA box region accessible to binding factors. These sites can be identified as being sensitive to attack by DNAase (Eissenberg et al.,1985).

CHAPTER 4: TRANSCRIPTIONAL ACTIVATOR PROTEINS.

Introduction.

The mechanism by which a protein bound to a site many base pairs from a gene activates transcription remains unclear. The most likely theory involves an activator protein which binds to cis-acting elements upstream from the gene and interacts with another protein. This secondary protein can be either bound to an initiation complex which could include the polymerase or can facilitate the binding of the polymerase to the DNA in the region of the transcription start site. The intervening DNA loops out to facilitate the reaction (Ptashne, 1986; 1988). The cis-acting sites could for example be enhancers as found in mammalian cells. In lower eukaryotes such as yeast, elements called upstream activator elements (UAEs) are required for efficient transcription. When analysed using reverse genetic techniques they function independent of orientation and are effective at some distance from the gene. Like mammalian enhancers they are the sites for the binding of specific trans-acting factors for example Gal4 and GCN4 found in yeast.

DNA binding proteins and transcriptional activation.

A comparison of different DNA binding transcription factors indicates that several types of protein domains have been adopted to mediate sequence-specific protein-DNA interactions. One prevalent structural framework for binding DNA is the so called "zinc-finger"

motif, which has been found in a large number of established and putative transcription factors (Klug & Rhodes, 1987). Another DNA binding structure that has been identified is the homeodomain, first recognised in homeotic genes responsible for regulating Drosophila development but now also found in a variety of other transcriptional control proteins from yeast to man (Levine & Hoey, 1988). In general, the binding of these domains to their DNA recognition elements is not sufficient to trigger transcription, additional protein sequences are necessary for transcriptional activation. These protein domains are thought to interact with RNA polymerase II or other ancillary factors of the transcription complex. One class of activation sequence is characterised by a high content of negatively charged amino acid residues that may be present in an amphipathic alpha helical structure (Ptashne, 1988). However, it has become apparent that there may be multiple types of transcriptional activating sequences. Recent evidence indicates that the mammalian DNA binding protein Sp1 activates transcription through a distinct motif, characterised by a high glutamine content (Courey & Tjian, 1988). These studies suggest that a number of different structural designs can be used for transcriptional activation, as in the case of DNA binding. The DNA binding and transcriptional activation domains of activator proteins are clearly seperable as shown for example in domain swap experiments on the yeast proteins Gal4 and GCN4 (Keegan et al., 1986).

DNA binding domains.

1) Helix-turn-helix.

Crystallographic studies determined for the coliphage 434 repressor, the cro protein of bacteriophage lambda and the CAP protein of E. coli, have revealed the existence of conserved α helical structures which computer modelling studies suggest are the DNA binding sites for these proteins (Laughon & Scott, 1984; Pabo & Sauer, 1984). The DNA binding structure comprises of two α helices connected by a β turn. Most of the DNA contacts made by cro, repressor and CAP occur on one side of the double helix. However, lambda repressor contacts both sides of the double helix by using a flexible region of protein to wrap around the DNA (Pabo & Sauer, 1984).

Many other prokaryotic and some eukaryotic DNA binding proteins share a motif similar to the helix-turn-helix. Prominent among these are the homeobox containing proteins. The homeodomain is present in the products of early developmental genes of Drosophila and other species and its sequence is highly conserved in otherwise unrelated proteins. That DNA binding specificity is a property of the homeobox has been shown by homeobox "swap" experiments (Hoey & Levine, 1988; Levine & Hoey, 1988) and the evidence that point mutation and inframe deletions within the homeobox can completely abolish binding functions (Desplan et al., 1988; Levine & Hoey, 1988). This evidence supports the idea that many of the developmental genes encode transcriptional regulators.

According to the current model based on molecular and genetic data and on the crystal structures of the lambda and 434 repressors and the CRO protein, the N-terminal part of the recognition helix fits into the major groove of DNA. Amino acids on one face of this helix establish hydrogen bonds and van der Waals contacts with the base pairs. The specificity-determining residues occupy positions 1 and 2 in the first turn of the recognition helix and positions 5 and 6 in the second turn (Berg, 1986; Anderson et al., 1987; Wolberger et al., 1988). An adjacent α helical region contacts the DNA backbone and may help to orientate the "recognition" helices.

The homeodomain was initially expected to bind to DNA in a conformation similar to that described for prokaryotic proteins. Indeed, homeodomains of such proteins as ftz (fushi tarazu), which have very similar recognition helices to these prokaryotic proteins, can recognise the same sequences (Desplan et al., 1988). This is consistent with the idea that the helix-turn-helix is important in generating DNA specificity. Homeodomains from more diverged genes for example prd (paired) do not bind to the same recognition sequence (Hoey & Levine, 1988). Treisman et al., (1989) demonstrate that the specificity of this homeodomain is critically dependent on the nature of the amino acid at position 9 of its recognition helix. This observation was also observed for the bicoid activator protein (Hanes & Brent, 1989) supporting the model that this amino acid is the major determinant of the specificity of bona fide homeodomain proteins. Studies on the homeobox protein paired

(Treisman et al., 1989) also suggest that in addition to the specificity directed by residue 9, individual homeodomain proteins may have other complexities in their interaction with DNA. The region just C-terminal to position 9, which is the most highly conserved part of the homeodomain, contains a stretch of 4 lysine or arginine residues. This region forms a helical C extension of the recognition helix (Otting et al., 1988). It is possible that this very basic region can interact closely with the phosphates of the DNA backbone, tilting the recognition helix to bring the amino acid at position 9 closer to the major groove than in the 434 repressor. This could provide an explanation for the conservation throughout evolution of the differences between the various classes of homeodomains outside the recognition helix.

2) Zinc-finger domains.

A number of DNA binding domains that are clearly different from the helix-turn-helix motif have been discovered. Prominent among these are domains that are structurally organised around zinc ions coordinated to invariant cysteines and/or histidine residues (Evans & Hollenberg, 1988). These domains are collectively referred to as "zinc-finger" domains. However, more detailed analysis clearly indicates that there are several distinct classes of domains that share the property of containing zinc ions but need not share additional structural characteristics:

i) Transcription factor IIIA-type proteins.

Originally discovered in Xenopus laevis, transcription factor TFIIIA typifies this type of activator protein. They contain a finger motif tandemly repeated a minimum of two times with a 7-8 amino acid linker seperating the units. Each finger unit contains a pair of cysteine and histidine residues separated by a loop of approximately 12 amino acids (Evans & Hollenberg, 1988). Such fingers are commonly referred to as C_2H_2 type. This module can be strung together to form proteins that can recognise long stretches of DNA sequence. The repetitive nature of the fingers suggests a structure in which amino acids at the tip of each finger loop interact directly with the DNA helix. For example, each of the 9 fingers of TFIIIA have been shown to interact with approximately 5 nucleotides (Fairall et al., 1986) with adjacent fingers interacting with contiguous stretches of DNA. In contrast Sp1 which has 3 fingers binds only to a 10 bp nonpalindromic GC box sequence rather than to the expected 15 bp, but with high affinity. Structural characterisations of these domains based on the predictions from amino acid sequence (Berg, 1988) has revealed a structure that includes an α helix that could lie in the major groove of the DNA to make specific contacts, analogous to those of the helix-turn-helix proteins. The precise correlation between finger number, nucleotides bound and affinity remains to be elucidated.

ii) The steroid receptor type proteins.

These proteins have non-repetitive fingers and have a variable number of conserved cysteines available for metal chelation. They include the yeast transcriptional factor Gal4 which has 1 finger incorporating 6 cysteines, and steroid receptors. Steroid receptors contain 2 apparently unrelated fingers coded for by separate exons, containing 4 and 5 cysteines respectively (Evans & Hollenberg, 1988; Berg, 1989). Each finger is separated by approximately 15-17 amino acids. The binding sites for this class of protein, collectively referred to as hormone responsive elements (HRE's), are relatively short and contain dyad symmetry implicating dimer formation for DNA binding (Berg, 1989). Recently Umesono & Evans (1989) showed that two distinct regions outside the loops of the two zinc fingers are critical for sequence recognition between the glucocorticoid, oestrogen and thyroid receptor elements and that this implicated region may form a recognition helix that is potentially analogous to those in the helix-turn-helix and TFIIIA type proteins. Further evidence (Berg, 1989) suggests that the first zinc finger and critical adjacent residues of steroid receptors are involved in protein-DNA interactions whereas the second zinc finger may be involved only in protein-protein interactions. In addition the binding specificity of steroid receptors can be altered by simple amino acid changes in the first zinc finger.

DNA activating domains.

The activating domains of many DNA activating proteins including the glucocorticoid receptor (Godowski et al., 1988; Hollenberg & Evans, 1988) the yeast activator Gal4 (Ma & Ptashne, 1987) and GCN4 (Hope & Struhl, 1986) and the HSV activator VP16 (Sadowski et al., 1988) have a highly acidic character produced by a propensity of negatively charged amino acids with an ill-defined structure. Ma & Ptashne (1987) found that moderate transcriptional activation could be achieved by fusing the Gal4 DNA binding domain to random polypeptides encoded by fragments of genomic E. coli DNA. Activating sequences were shown to contain an excess of acidic side chains with the strength of activation related roughly to the amount of negative charge.

The role of these unstructured activating domains in the formation of transcriptional activation complexes may be determined from the structure of the eukaryotic RNA polymerase II (RNA Pol II). It has been suggested that the long heptapeptide repeat tail at the carboxy terminus of the polymerase II large subunit might contact sequence specific transcription factors (Allison et al., 1985; Zehring et al., 1988) perhaps through hydrogen bonding (Sigler, 1988). Nearly all the amino acid side chains of the repeating blocks are hydrophilic (primarily hydroxyl groups and never acidic) and this structure is highly conserved among many eukaryotes and prokaryotes (Allison et al., 1988). The unique secondary structure proposed for this domain is one which projects out of the globular polymerase

polypeptide. The cleavage of repeating elements from the catalytic domain of RNA Pol II is consistent with this observation (Zehring et al., 1988). The RNA polymerase is postulated to interact with one or more activating acidic regions by hydrogen bonding between the carboxylate groups of the activating domain and the hydroxyl groups of the RNA Pol II subunit. The "non-specific" nature of the interaction means that RNA Pol II can function in a wide variety of different transcriptional-activation systems.

Once RNA Pol II has been positioned accurately on the template it must be released from the initiation complex to proceed with transcription. The above model suggests that this could be accomplished by phosphorylation of the hydroxyl groups at or near the interaction sites of the activator domains. This would break the hydrogen bond contacts and may concurrently activate the transcriptional function of the enzyme (Sigler, 1988).

ii) Non acidic activating domains.

Analysis of Sp1 mutants reveals multiple distinct regions outside of the DNA binding domain that are responsible for mediating transcriptional activation (Courey & Tjian, 1988). The two most active domains consist of unusual structure with a very low charge density, but a strikingly high glutamine content. A number of other known and suspected transcription factors also contain glutamine-rich segments for example the Drosophila zeste gene which has been shown to bind the ultrabithorax promoter (Biggins et

al., 1988) and the products of a number of homeobox containing genes such as Antp (Schneuwily et al., 1986) and Cut (Blochlinger et al., 1988). The amide moieties of the glutamine side chains can be envisaged to participate in hydrogen bonding to the RNA polymerase or other activating proteins. Glutamine may be unique in being able to reach a contact point in a depression or groove of another protein which would be inaccessible to other residues due to structural constraints. In addition the activating regions of the human oestrogen receptor does not contain any stretches of acidic amino acids (Webster et al., 1989; Tora et al., 1989). Thus it appears that acidic regions are not the only transcriptional activation motif. Tora et al., 1989 propose that different activating domains may interact with different components of the basic initiation complex whose formation requires several ancillary proteins in addition to RNA polymerase II (Buratowski et al., 1988).

Summary: Trans-Activating Proteins

The function of trans-activating proteins is to stimulate the formation and/or action of transcriptional preinitiation complexes. The core of these complexes consist of the promoter-binding proteins, e.g. TFIIIA, on which the RNA polymerase depends for correct placement and orientation with respect to the start site of transcription. The activating proteins are targeted to specific DNA sequences for example HRE's usually by a specific DNA binding domain on the same polypeptide.

Ligand binding receptors.

One group of proteins engage in the initiation of transcription from certain genes only in response to the specific binding of ligands to them. They include factors which can bind steroids, thyroxine, retinoic acid and vitamin D. The structural homology observed in these proteins suggests that they have all evolved from a common ancestral gene (Evans, 1988). Ligand binding receptors have three functional domains, the DNA binding domain, the activating domain and a unique ligand binding domain.

The ligand binding domain is characterised by a large number of hydrophobic amino acids. The binding of a ligand to this site results in an activation of the protein to bind DNA. Control of the glucocorticoid receptor is governed by the protein HSP90 which binds

to it and prevents it from entering the nucleus (Denis et al., 1988 and references therein). The hormone dexamethasone dissociates the complex and allows the receptor to move into the nucleus and activate transcription by uncovering specific DNA binding sites. In the presence of hormone, binding of the ligand appears to involve the dimerization of the receptor which is necessary for the activation of transcription (Green & Chambon, 1988). The oestrogen receptor is also believed to be inhibited by HSP90 (Sabbah et al., 1987).

The major characteristics of a hypothetical nuclear activating protein are summarised in Figure 3.

Control of gene activator proteins by Phorbol Esters: The Ap-1 transcription factor.

The expression of some genes is influenced by external chemical stimuli for example phorbol esters. Phorbol esters, such as 12-O-tetradecanoyl-phorbol-13- acetate (TPA) are potent tumour inducers. Among the genes whose expression is altered by TPA are the cellular proto-oncogenes c-fos, c-myc, c-sis and collagenase. Angel et al., (1987) show that promoters of TPA-inducible genes contain a common cis element which can confer TPA inducibility on heterologous promoters via a TPA modulated trans-acting factor. These cis-acting elements also bind the cellular protein Ap-1. Lee et al., (1987) demonstrated that multiple Ap-1 binding sites can confer TPA inducibility in various plasmid constructs after

Figure 3.

A summary of the characteristic features of a Nuclear Activating Protein and its hypothetical structure.

A) Ligand binding domain.

The binding of a ligand (X) (eg dexamethasone) to this site may activate the protein receptor by the displacement of an inhibitory protein (Y) (eg HSP90).

B) DNA binding domain.

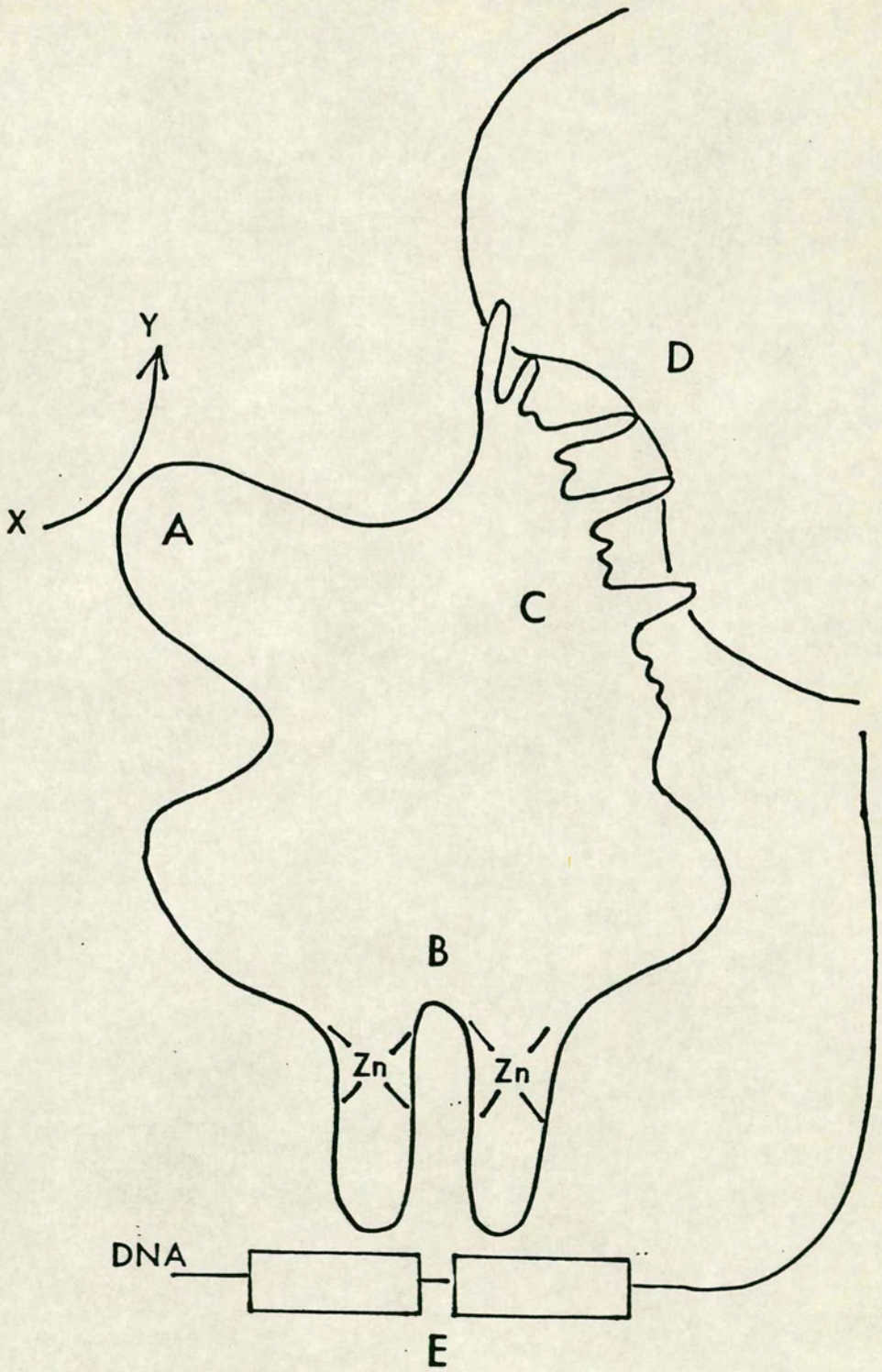
The region is well structured and characterised by projections of amino acids anchored by zinc. Multiple copies of the zinc fingers increase DNA interactions to give multiple DNA binding specificities and bring distant DNA sequences together through their interactions.

C) Activating domain.

A poorly structured acidic region with a predominance of negatively charged amino acids. These have the potential to interact with the protein/polymerase complex (D) by the formation of an amphipathetic alpha helix.

E) Specific DNA sequence motifs.

These are recognised by specific nuclear activator proteins (eg Sp-1, Ap-1, hormone receptors, Gal4).


Y
X
A
D
C
B
Zn
Zn
DNA
E

transfection into HeLa cells. A point mutation which abolishes the *cis* element activity also strongly disrupts Ap-1 binding. This suggests that the induction of genes via phorbol esters is mediated by the binding of Ap-1 to specific DNA elements. Ap-1 may also have a role in establishing a basal level of transcription in conjunction with other cellular proteins for example with Sp1 in the SV40 (see Figure 2A) and human metallothionein promoter (Lee *et al.*, 1987b). Phorbol esters are considered to exert their effect on gene expression in a pathway involving the activation of protein kinase C (Nishizuka, 1984). This enzyme may regulate gene transcription by the phosphorylation of Ap-1 itself or of an inhibitory protein which regulates its activity (Angel *et al.*, 1987). Corssi *et al.*, (1988) have shown that *c-fos* induction by phorbol esters involves a *trans*-acting serum responsive element in a pathway involving adenylate cyclase.

Evolution of transcription factors.

The large subunit of RNA polymerase B is conserved between yeast and man. Recent research has found that upstream regulatory factors also exhibit interspecies conservation. Kakidani & Ptashne (1988) show that the yeast transcriptional activator Gal4 can activate a TATA box containing promoter in HeLa cells. The oestrogen receptor has been shown to activate a similar promoter in yeast (Webster 1988; Metzger *et al.*, 1988). In addition a factor which can functionally substitute for the mammalian TATA box factor (TF11) has been found in yeast (Cavallini *et al.*, 1988; Buratowski *et al.*, 1988). The

products of the human proto-oncogenes c-fos and c-myc or oncogenes v-fos and v-myc can activate transcription in yeast when fused to the DNA binding portion of the LexA protein (Lech et al., 1988). The JUN oncoprotein, which causes sarcoma in chickens, bears a region of extensive homology with the DNA binding portion of the yeast GCN4 protein. This suggests that JUN may bind to DNA in a sequence specific manner, recognising the same sequences as GCN4 (Struhl, 1987). The avian JUN protein and the c-fos protein (Rausher et al., 1988) have a similar DNA binding specificity to the mammalian Ap-1 protein implying that they may be derived from Ap-1. The thyroid receptor also shows extensive homology to the viral oncogene erbA (Evans, 1988 and references therein). These results together indicate an ancestry of yeast, avian and mammalian transcription factors. More importantly it implies that the oncogenic state in some cases may be due to the altered expression of genes coding for activator proteins.

Activator protein interactions.

DNA binding of transcriptional activator proteins brings to the DNA in the vicinity of the gene an activating region which interacts with some component of the transcriptional machinery. This interaction promotes transcription. The activating region might also interact with the target protein freely in the nucleus when neither is bound to the DNA. The activator proteins, which differ only in the specificity for a DNA binding site, can compete for the same target protein. If present in large concentrations, those activators

which lack the necessary DNA binding sites may titrate out all of the target proteins. This would lead to inactivation of the gene whose activating molecules bear the binding site. This hypothesis has been termed "squelching" and has been used to explain why strong activators such as those produced during viral infections are only expressed transiently (reviewed in Ptashne, 1988). To overcome this otherwise detrimental effect to the cell it has been supposed that cell activators may work cooperatively with one another. A very strong activating region is produced when a number of cell activators are bound to the DNA whilst each one in isolation is too weak to have much effect on binding the target protein when unbound. In such a system any two eukaryotic activator proteins can work synergistically because they simultaneously make contact with the target protein (Kakidani & Ptashne, 1988; Schaffner, 1989). The typical gene, however, bears in its vicinity an array of distinct binding sites for regulatory proteins (Maniatis et al., 1987). Attaching these sites to a second gene in place of its ordinary regulatory elements does not inactive the gene, but rather confers on it the pattern of expression of the first gene. From this evidence it seems that the total number of activators is small compared with the number of genes in higher eukaryotes. Tissue-specific expression is accomplished by the interaction of different combinations of transcription factors with physiological stimuli. A target protein may become attached to one or more protein activators which facilitates the binding of the polymerase. Any activators which simultaneously touch the target protein will work synergistically and additional DNA bound activators will add their effects linearly.

CHAPTER 5.

DNA structure of the Mup BS2 promoter.

On the basis of nucleic acid hybridisation experiments, the Mup genes can be subdivided into two main groups group 1 and group 2. Each group has about 15 members. A small number of other genes exist which belong to other groups (see Chapter 1, Bishop et al., 1982). The full sequence of the transcription unit of a group 1 gene BS6, and a group 2 gene BS2, have been determined along with about 700 bp of flanking sequence (Clark et al., 1985b). While the available evidence indicates that all group 1 genes are true genes, all of the group 2 genes so far examined are putative pseudogenes. BS2 carries three lesions in its protein coding sequence and could not be translated to yield a protein with the molecular weight of MUP (Clark et al., 1985b). Studies with a synthetic oligonucleotide probe, specific for BS2, failed to pick up any transcipts from this gene (McIntosh & Bishop, 1989; Shahan et al., 1987b).

It does not necessarily follow from these observations that the promoter sequences of the group 2 Mup genes are redundant in their capacity to initiate transcription from the putative BS2 CAP site. Sequence analysis of both BS6 and BS2 show that both genes possess all the sequences recognised to be essential for this event (see Chapter 3). Alignment of the BS6 and BS2 sequences to maximise homology show that both genes possess a TATA consensus sequence at position -31 bp. The CAP site of BS6 has been accurately mapped by primer extension and SI nuclease protection experiments (Clark et

al., 1985b) and sequence homology has placed the BS2 CAP site at the same position. In addition both genes contain identical CCCAT sequences at position -109 bp (BS6) and -77 bp (BS2). The two gene sequences show an 11.1% divergence over the 5' flanking region and a divergence of 13.4% for the full sequence (Clark et al., 1985b).

The BS6 and BS2 promoters contain important sequence differences.

The $C(A)_N$ region.

The BS6 and BS2 promoters contain sequence differences which may contribute to the transcriptional and translational efficiency of each promoter and the observed patterns of tissue-specific expression. The most prominent differences in sequence occur within an adenine (A) rich region just 5' of the TATA box from positions -47 to -93 bp. This region contains primarily A or $C(A)_N$ nucleotides and varies in length in the different clones from 15 (BS2) to 46 (BS6) nucleotides (see Figure 4A; Clark et al., 1985b; Held et al., 1987). Tissue-specific expression studies in transgenic mice show that this region is most likely concerned with the strength of the promoter rather than governing tissue specificity (Al Shawi et al., 1988 & unpublished results). Repeats of this nature have been observed in other genes. The 5' flanking region of genes encoding mouse complement C4 and the related testosterone regulated sex-limited protein also have a high degree of overall homology (95%) (Nonaka et al., 1986). However, the sex-limited protein gene lacks 31 and a 60 nucleotide segments

Figure 4.

4A) Sequence alignment of the A-tract region located from -40 to -90 bp upstream of the Mup genes initiation site.

Group 1 genes: BS1, BS6, BS109-1, BS5, BL7, BL1, CL8, CL11.

Group 2 genes: BS2, BS102-2, BS109-2, BL25.

Rat gene, α-2u globulin: A2u.91.

4B) To show the DNA sequence signals present in BS6 and BS2.

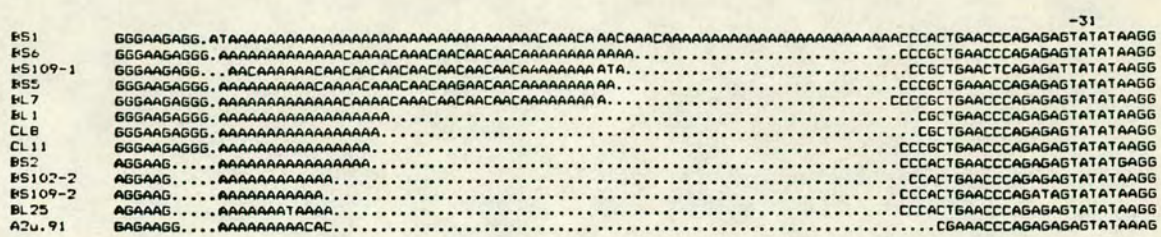
Consensus sequences for the transcription initiation * are taken from Breathnach & Chambon (1981) and for the translation initiation ** from Kozak (1984a).

SIGNAL	CONSENSUS	BS6	BS2
Transcription	*GGYCAATCT	GACCCATAC	GACCCATAC
Initiation.	(-80)	(-109)	(-77)
	GNGTATAWAWNG	GAGTATATAAGG	GAGTATATGAGG
	(-30)	(-31)	(-31)
Translation	**CCRCCATG	CCAAAATG	ACCAAATG
Initiation.		(+67)	(+66)
Translation		Exon 6: TGA	Exon 1: TGA
Termination.		(+2854 bp)	Exon 3: TAA
			Exon 3: TGA
			(+156, +1422, +1482 bp)

4C) Regulatory sequence motifs found within the Mup promoter.

The approximate positions of these motifs refer to the BS6 and the BS2 sequences. The reverse strand of the consensus is denoted "rev" and the lower case letters refer to mismatches with the consensus sequence. Subscripts below the motifs indicate those genes which differ from the sequence shown. (Ghazal, 1986).

A.



C.

FACTOR/REGULATORY SEQUENCE	CONSENSUS	SITES	GROUP 1 GENES	APPROX POSITION	GROUP 2 GENES	APPROX POSITION
Nuclear Factor 1 (NF1)	TGGN$_{6-7}$KYCAM	a	TTGcAN$_7$CCA	-305	GTGcCN$_7$CCA	-270
	rev)KTGRMN$_{6-7}$CCA	b			GTGAtN$_7$CCA	-225
		c	TGGN$_{6}$aTCAA	-168	TGGN$_{6}$aTCAA	-135
		d	TGGN$_{6}$GCCAC	-1		
		e			TGGN$_7$GCCAC	+240
		f	TTGAAN$_6$CCA	+285	TTGAAN$_6$CCA	+315
Metal Responsive Element (MRE)	CYTTTGCRYYCG	1			aGGGatCAAAAG	-710
	rev)CGRRYGCAAARG	2			aGG*ATGCccAGG	-660
		3	CCgcTGCACCaG	-515	a in BS109-2	
	(Palindrome)	4	CaGATcC*AAAGt	-315	CCaaTGC*A*T*T*aG	-480
			g in BS1, BL1		aaaa in BL25	
		5	aGttTGC*AAAGG	-195	CaGATcCAAAGG	-280
			t in BS109-1			
		6			CTAATGCAtAGg	+430
Enhancer Core Sequence (E)	GTGGWWWG	1			CTATCCcC	-830
	rev)CWWWCCAC	2	CAAACCAg	-320		
		3	CAATCCAt	-220	CAATCCAt	-185
		4	CATTCCAC	-140	tATTCCAC	-110
		5	CAAACaAC	-75		
			present only in BS1, BL7 & BS109-1			
Glucocorticoid Responsive Element (GRE)	BKKYWCWWYSTGTYCT	1			tTGaACc-TGTGTCCT	-535
	rev)AGRACAGRWKCWRWWG	2	CTaTA*CAtACtTGTCCT	-380	CTaT*A*CAtACtTGTCCT	-340
			g in BS1		aa in BL25	
		3	CaGTTCAgTCTGTaCT	-360		
		4	AGGACAaGcaaAGggG	-25	AGGACAaGcaaA*GgAG	-25
					g in BL25	
CCAAT box	GGYCAATCT		GaCCcATaCa	-113	GaCCcATaCa	-81
Globin genes conserved -100 region 23-25bp upstream of CCAAT box	RRCYYCACCC		GAaaaCAttCC	-147	GAaaatAttCC	-105
Sp1	KGGGCGGRRY		NONE		NONE	
Heat Shock Element (HSE)	CTNGAANNTTCNAG		NONE		NONE	

containing ACACCC and ACAC repeats respectively, suggesting that simple repetitive sequences may play some role in differential regulation of these genes. Also, $C(A)_N$ repeats or A-rich regions have been implicated in the formation of bent DNA in trypanosome kinetoplast DNA (Wu & Crothers, 1984) and at a Saccharomyces cerevisiae autonomously replicating sequence (Snyder et al., 1986) and may play some regulatory role.

The TATA box sequence of BS6 and BS2 compared.

Another feature which discriminates between the two genes is an A to G transversion at position 6 in the TATA box sequence of BS2 (TATATAA (BS6): TATATGA (BS2)). The BS6 sequence conforms with the consensus of Breathnach and Chambon (1981) derived from 50 different genes (Figure 4B). However, none of these genes possessed a G in the 6th position. This transversion may affect the strength of the BS2 promoter or totally abolish its function of fixing the initiation of transcription at the CAP site. Interestingly the feline c-myc gene (Stewart et al., 1986) and ovine β-lactoglobulin gene (Ali & Clark, 1988) exhibit the same G nucleotide at position 6. The TATA box sequences are TATATGC and TATAAGG respectively. Neither of these sequences affect the transcription of the gene. Primer extension studies on both of these RNAs have unambiguously defined the 5' terminal sequences which constitute the CAP sites approximately 30 bp downstream of the two TATA boxes.

Other differences between the BS6 and BS2 sequences include: i) a small duplication of 2 bp in the slice acceptor site of intron 6

that may favour the formation of observed short form mRNA, ii) An alteration in the translation initiation signal CCAAA to ACCAA that may impair translational efficiency (Figure 4B) and iii) An inframe increase in the length of the signal peptide region.

Transcriptional activator protein binding sites in the Mup promoter.

The likelihood that tissue specific factors may be required for Mup promoter activity is strengthened by an examination of its sequence. About 700 bp of 5' flanking sequence has been shown to contain a series of sites that could serve as binding sites for trans-acting tissue-specific/regulatory factors. These are the Nuclear Factor 1 (NF1) sites, BS6 (4 copies) and BS2 (5 copies). Multiple copies of the enhancer core sequence BS6(4), BS2(3), the glucocorticoid response element (3 copies in each) and the metallothionein gene metal responsive element BS6(3), BS2(5) are also found. However, only the glucocorticoid regulatory element occurs at a frequency significantly different from random expectation (Ghazal, 1986). One of the putative metal response elements lies within a 25 nucleotide palindrome present in BS6 but not in BS2 genes. The palindrome exhibits some homology with the metallothionein metal response palindrome (Figure 4C). Sequence analysis indicates that Mup genes have an Ap2 binding site sequence at about -500bp with a match of consensus of 7 out of 8 nucleotides (Held et al., 1987).

To determine whether the BS2 promoter is able to initiate faithful transcription *in vivo*, standard reverse genetic techniques were implemented.

Mup expression in cell lines.

Mup genes are expressed primarily in the liver (Clissold & Bishop, 1982). Liver tissue is appropriate for studying the expression of differentiated function because it is composed predominantly of a single type of differentiated cell, the hepatocyte, and hepatocytes express a large number of well characterised proteins including transferrin, albumin and alpha fetoprotein. Systems available for studying the regulation of liver-specific genes include liver from intact animals, liver slices, primary hepatocytes in culture and cocultures of primary hepatocytes with epithelial cells. Primary hepatocytes in culture are useful for these studies because they can be derived from normal adult liver, and when maintained under the appropriate conditions they produce levels of albumin mRNA similar to that produced by the liver. The disadvantage of primary hepatocytes is that unlike a cell line they do not replicate, making it difficult to readily obtain large quantities of cells. Cultures of differentiated primary hepatocytes also require the use of collagen or a matrix. Cocultivation with epithelial cells eliminates this problem but results can be complicated by the effects caused by the presence of the second cell type. Work has been carried out on the expression of *Mup* genes in hepatocytes (Clayton & Darnell, 1983; Spiegelberg & Bishop, 1987). When hepatocytes are plated the expression of *Mup* genes is rapidly

switched off. This results in a dramatic fall in the level of Mup mRNA to virtually zero in about 4 days. During the same time period the level of transferrin mRNA doubles and alphafeto protein mRNA appears denovo. This shows that the integrity of the cells is maintained in culture during changes in expression which seem to mimic liver regeneration. "Run-on" experiments with isolated nuclei showed that the fall in the level of Mup mRNA could be explained by a decrease in the rate of RNA transcription from the Mup genes. It seems possible that these changes might not occur under different culture conditions. Spiegelberg & Bishop (1987) show that growth hormone, thyroxine and insulin all retard the the rate of decay of Mup transcription in cultured hepatocytes. Growth hormone and thyroxine also show a synergistic effect.

Liver derived cell lines, including hepatoma cells and immortalised hepatocytes, have been used to study the expression of liver specific proteins. Hepatoma cell lines replicate continuously in culture without a matrix. Although hepatoma cell lines are derived from tumour and not normal tissue, the H411EC3 rat hepatoma (Ohanian et al., 1969) and the HepG2 human hepatoma (Knowles et al., 1980) cell lines have been used succesfully in studying liver specific proteins. One limitation to the use of many hepatoma cell lines is that they are grown in serum which contains undefined substances that can alter gene expression. In general, a disadvantage of most hepatoma cell lines is that they express liver specific genes at much lower levels than those expressed by the intact liver. A further complication is the fact that different immortalised cell lines from a given tissue show considerable

variability in their phenotypic expression. Recently workers have derived a series of SV40 immortalised rat hepatocyte cell lines which produce albumin and albumin mRNA at levels found in the normal liver and freshly isolated hepatocytes (Woodworth & Isom, 1987). It is not known whether these cell lines express α2u-globulin.

Mup genes can be expressed in BHK fibroblasts when coupled to an SV40 enhancer.

Held et al., (1987) transfected mouse L cells with Mup genomic clones but only observed a low level of expression. High level Mup gene expression may well require multiple tissue and hormone specific factors which are not present in L cells. Ideally the study of Mup expression would have been carried out in a mouse liver cell line which expresses the endogenous Mup genes. To date there are no such lines available. Instead attention was focused on the fact that although a large number of promoters work very weakly or not at all in some cell lines, their activity can be forced to higher levels by linking enhancer sequences to them in cis. Examples are the β- globin (Banerji et al., 1981), conalbumin (Wasylk et al., 1983) and lysozyme promoters (Renkawitz et al., 1984). The SV40 enhancer, being one of the strongest enhancers and active in many cell types, is often chosen for this purpose. It was found that heterologous coding sequences linked to the promoter of the group 1 Mup gene BS6, could be expressed in Baby Hamster Kidney (BHK) fibroblast cells when coupled to the SV40 enhancer, but not in its absence (Bishop, unpublished). This establishes that the BS6 promoter can be activated but reveals nothing about its

behaviour in the liver. Its inability to function as a promoter on its own in fibroblasts is of course consistent with the lack of expression of endogenous mouse Mup genes in these cells. S1 nuclease and primer extension assays show that RNA messages were being faithfully initiated at the Mup CAP site (Ghazal, 1986). The BS6 promoter showed a unique property in that it abolished the promoter activity of the upstream SV40 promoter, although the latter is closer to the SV40 enhancer (Ghazal, 1986).

Constructs were made to test the fidelity of the BS2 promoter in BHK fibroblasts. The basic components of the plasmid were an SV40 enhancer and early promoter and 2.2 kb of 5' flanking sequence from the group 2 Mup gene BS2, linked to the coding region for the Herpes simplex virus thymidine kinase type 1 gene. To facilitate the rescue of the hybrid gene from the genomic DNA of transgenic mice the construct contains the bacterial amber suppressor tRNA gene SupF (Goldfarb et al., 1982).

Viral thymidine kinase and bacterial neomycin resistance genes in transfection systems.

Thymidine kinase is an enzyme which catalyses the phosphorylation of thymidine to thymidine monophosphate in the salvage pathway of pyrimidine synthesis. Herpes simplex virus encodes a novel thymidine kinase. The transcribed portion of the gene is approximately 1300 nucleotides in length and contains no intervening sequences (Wagner et al., 1981). There is an untranslated region of 107 nucleotides at the 5'end of the mRNA followed by an open reading frame of 1128 nucleotides which codes for 376 amino acids. The promoter region of the Herpes simplex virus (HSV) thymidine kinase type 1 gene (HSV*tk*) has been extensively analysed and shown to contain the TATA and CCAAT consensus signals and two GC rich repeated sequences in the 105 bp upstream from the normal CAP site (McKnight & Kingsbury, 1982; McKnight & Tjian, 1986). See Figure 2B. The availability of this information enables the separation of the *tk* gene into its promoter and coding region fragments. This can be achieved by digesting the gene with *Hinc*II or *Bgl*II, both of which cut once within the region corresponding to the 5' untranslated region of the mRNA. The 5' fragment can be used as a promoter for DNA sequences which lack a functional promoter, but which contain all the other necessary components for example the initiation codon. A deletion of the promoter region reduces the number of thymidine kinase proficient transfectants to 0.1-1% relative to an intact *tk* gene (McKnight & Gavis, 1980). The substitution of this region by exogenous DNA sequences has allowed the development of a suitable assay system for the identification

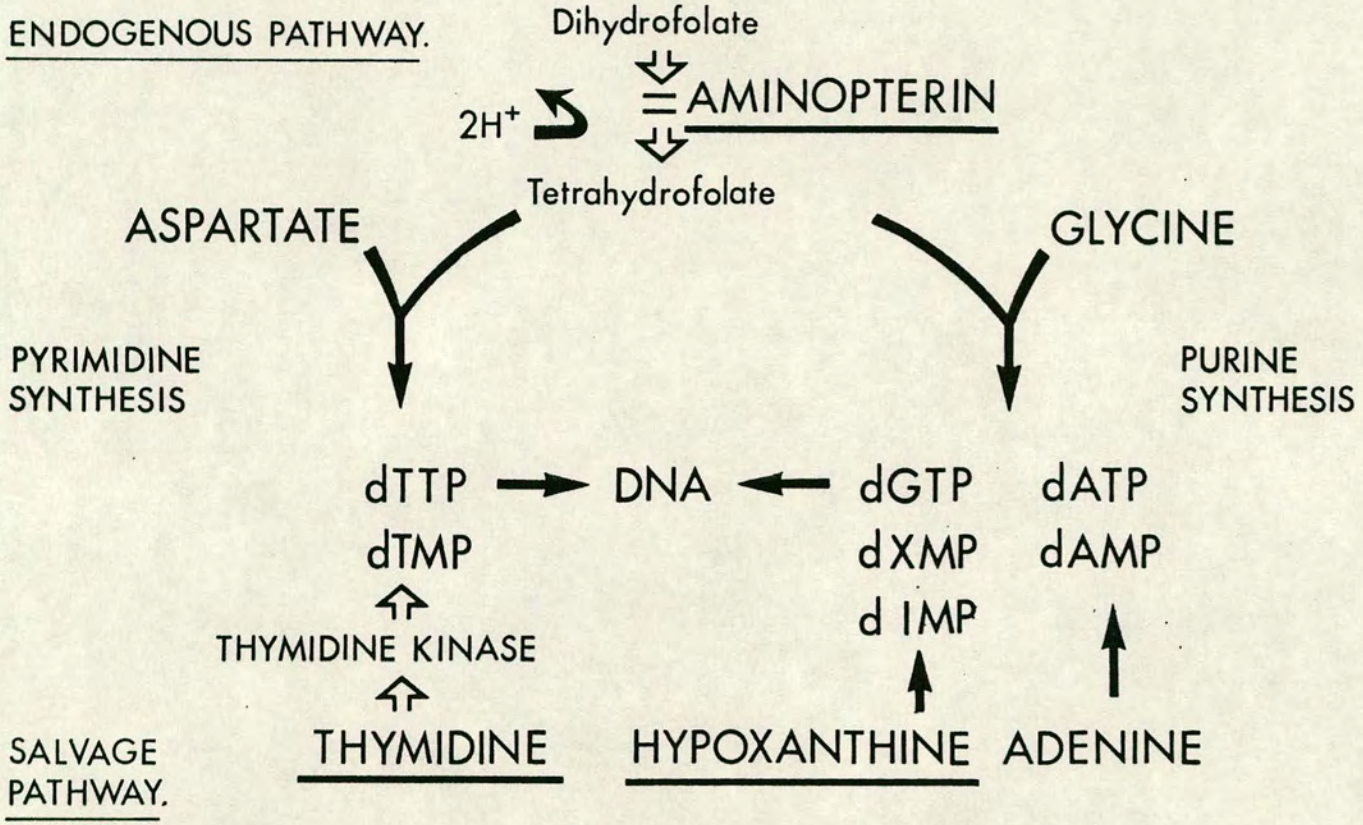
and analysis of eukaryotic control elements. The promoter deleted *tk* gene can be convieniently isolated in a 1759 bp *Bgl*II-*Bst*EII fragment containing 57 bp of leader sequence (Wilkie *et al.*, 1979). It is inactive in gene expression studies but can be reactivated by insertion of DNA sequences that contain *cis*-acting regulatory elements. The TK^+ phenotype can be efficiently selected over a TK^- background by utilising growth conditions in which the pyrimidine salvage pathway enzyme, thymidine kinase, is necessary for survival (Littlefield, 1964). The growth medium HAT contains Hypoxanthine, Aminopterin and Thymidine. Aminopterin inhibits the *de novo* synthesis of dATP, dGTP and dTTP from dihydrofolate. Hypoxanthine is a substrate for the salvage pathway for dATP and dGTP and therefore allows these nucleotides to be synthesised. Thymidine is essential for the synthesis of TTP and is totally dependent on an active *tk* gene (Figure 5). It has been suggested from experiments which show an amplification of truncated, promoterless *tk* genes in HAT medium that the survival of a cell in HAT medium may involve some threshold level of expression (Roberts & Axel, 1982). Any cell which expresses a *tk* gene below that threshold will not survive.

There exist cell lines deficient in TK with low rates of spontaneous reversion to the TK^+ phenotype which can be used as recipients for example Baby Hamster Kidney (BHK) fibroblast cells (Macpherson & Stoker, 1962). Under appropriate conditions either the TK^+ or TK^- phenotypes can be selected. TK^- cells are resistant to the thymidine analogue BUdR. The *tk* gene is thus an ideal subject for mutational analysis. The gene product, thymidine kinase, is a

Figure 5.

A diagram to show the biosynthetic pathways of dinucleotide triphosphates, dATP, dGTP, dTTP.

HAT medium contains hypoxanthine, aminopterin and thymidine. Aminopterin blocks the endogenous synthetic pathway of dinucleotide triphosphates and the salvage pathway is stimulated. This requires hypoxathine and thymidine as substrates. A functional thymidine kinase gene is essential for survival of a cell in HAT medium.


ENDOGENOUS PATHWAY.
Dihydrofolate
2H+
AMINOPTERIN
Tetrahydrofolate
ASPARTATE
GLYCINE
PYRIMIDINE SYNTHESIS
PURINE SYNTHESIS
dTTP
DNA
dGTP
dATP
dTMP
dXMP
dAMP
d IMP
THYMIDINE KINASE
SALVAGE PATHWAY.
THYMIDINE
HYPOXANTHINE
ADENINE

well characterised viral protein of known function which is distinguishable from the cellular enzyme in immunogenicity, molecular weight, substrate specificity and electrophoretic mobility. The active HSV*tk* gene can be detected in extracts of transfected cells by measuring the conversion of radiolabelled thymidine to thymidine phosphates. Since the virus encoded enzyme has different substrate specificities and feedback control from the cellular enzyme, cell encoded activity can be selectively inhibited by the addition of 0.2mM TTP allowing only the assay of HSV-thymidine kinase (HSVTK) (Jamieson & Subak-Sharpe, 1974). This property is utilised in the assays for tissue specific expression of the transgene in mice.

HSV thymidine kinase is representative of a type of selectable gene which can only be introduced into cells which are mutant and lack its expression. Other selectable genes can be introduced into any cell. An example is the bacterial neomycin resistance gene (*neo*). This gene, because of its bacterial origin, is normally nonfunctional in mammalian cells. However, the addition of a mammalian promoter and poly A sequences permit its expression. The *neo* gene codes for an aminoglycoside modifying enzyme which inactivates the aminoglycoside antibiotic G418. Mammalian cells are not sensitive to this antibiotic. The development of a drug, G418, which is cleaved by the *neo* gene product and to which mammalian cells are sensitive, enables selection of an active neomycin gene. G418 is usually used at a concentration of 400ug/ml. At lower concentrations (100ug/ml) there is a significant delay but the cells are eventually killed (Southern & Berg, 1982).

CHAPTER 6: MATERIALS AND METHODS.

Tissue culture methods.

The cells used in the experiments were baby hamster kidney fibroblasts which are deficient in thymidine kinase (BHK*tk*$^-$). The parent line of these cells was derived from the kidneys of 5 unsexed, one day old hamsters (Macpherson & Stoker, 1962). Cells were grown at 37^{o}C in 5 % CO_2 in Dulbeccos Modified Eagles Medium (DMEM) supplemented with 3.7gl^{-1}sodium hydrogen carbonate, 10% foetal calf serum (Gibco), penicillin (100u/ml) and streptomycin (100ug/ml; Gibco). Culture vessels were purchased from NUNC. Cells were detached by trypsinisation with 0.25% trypsin, 0.02% EDTA.

DNA mediated gene transfer.

Transformation experiments were carried out with exponentially growing recipient cells seeded 24 hours before the addition of donor DNA. The medium was not changed before the addition of the DNA solution. Transfections were carried out based on the calcium phosphate technique (Graham & van der Eb, 1973) described in the protocol by Spandidos and Wilkie (1984) with the modifications noted below. The DNA/calcium phosphate co-precipitate was maintained on the cells for 12 hours after which a pre-selection expression time in non-selective medium was established to give maximum expression of the transfected gene. The expression is maximal with pre-

selection expression periods of 36 hours for the BHK*tk*$^-$ cells used. After this period cells were either harvested for transient activity, or selected for HSVTK expression in DMEM +10% HAT (Gibco) (Hypoxathine, Aminopterin, Thymidine; Littlefield, 1964) or *neo* expression in DMEN + G418 (400 ugml^{-1}).

Long term stable expression. The DNA/calcium phosphate co-precipitate was added to the culture medium at a ratio of 0.7mls coprecipitate (containing 5ug of uncut Form 1 DNA) to 2 x10^5 cells, in 5mls of medium in a 4.5cm diameter dish . After 12 hours the medium was replaced with fresh non-selective medium for a further 36 hours, when selective medium containing HAT was applied. The medium was changed every 3 days for up to 14 days before the colonies were counted.

Transient Expression. The DNA/calcium phosphate co-precipitate was added to the culture medium at a ratio of 0.6mls coprecipitate (containing 15ug of DNA) to 8 x 10^5 cells in 10 mls of medium in a 9cm diameter dish. After 12 hours absorption and 36 hours pre-selection expression the cells were harvested and assayed for HSV thymidine kinase as described. Carrier DNA was omited from the transfections. Replica plates were carried out for each of the transfected recombinants.

Herpes simplex virus thymidine kinase assay.

The cells were harvested by washing twice in ice cold Phosphate

Buffered Saline (PBS) before they were scraped off into 5mls PBS and pelleted by centrifugation. A further washing with 5mls PBS and centrifugation followed, before the cells were taken up in 200ul of chilled sonication buffer containing 10mM potassium chloride, 2mM magnesium chloride, 10mM Tris, 1mM ATP, 10mM sodium flouride and 50mM aminocaproic acid pH 7.4 (Brinster et al., 1981). The cells were pelleted by centrifugation (10' at 15000g) after being lysed by sonication at 4°C for 5 hours. The enzyme assay is described by Spandidos and Wilkie (1984). The unit of thymidine kinase activity is the amount of enzyme which catalyses the formation of 1pmol of TMP in 1 minute. Protein was determined by the Bradford method (Bradford, 1976) for specific activity measurements. A 15ul aliquot of cell supernatant was added to 75ul of reaction mixture containing 10mM magnesium chloride, 150mM Tris (pH 7.5), 10mM ATP, 25mM sodium fluoride, 10mM β-mercaptoethanol and 15uCi of tritiated thymidine ^{3}H-TMP (90 Ci/mmol; Amersham). The mixture was incubated at 37°c. The ^{3}H-TMP produced was measured by spotting 25ul of mix onto DE-81 paper (Whatman) discs at time intervals 30, 60 and 120 minutes. The discs were washed 4 times, 10 minutes each, in 10mM Tris (pH 7.5) with shaking and dried under vacuum at 80°C. The radioactivity on the filters was determined by liquid scintillation counting (scintillant was 5g/l PPO, 0.3g/l dimethyl POPOP in toluene). The cell extracts were stored at -70°C. The assaying of thymidine kinase in mouse tissues was as described but with the following refinements. Tissues were dissected and immediately washed in ice cold PBS before being homogenised in sonication mix buffer. The reaction mix differed from that described by the addition of

0.4mM TTP which inhibited endogenous thymidine kinase by feedback control (Jamieson & Subak-Sharp, 1974).

Northern blots and DNA probes.

RNA was prepared from the cells by the guanidinium thiocyanate method (Chirgwin et al., 1979). After a PBS wash, 4mls of solution containing 4M guanidinium thiocyanate, 0.5% w/v sodium N-lauryl sarcosinate, 25mM sodium citrate, 0.1% w/v Sigma antifoam A and 0.1M β-mercaptoethanol, were added to a 9cm diameter plate containing 8 x 10^5 cells. The resulting solution was layered over 1.2mls of 5.7M caesium chloride buffered with 25mM sodium acetate (pH 5) and centrifuged at 36K, 20°C for 12 hours. The pelleted RNA was resuspended in 1ml of 7.5M guanidine hydrochloride, 25mM sodium citrate (pH 7), 5mM DTT and precipitated twice, once with 0.025 volumes of 1M acetic acid, 0.5 volumes of ethanol and once with 2mls double distilled water, 0.1 volumes of 2M sodium acetate and 2 volumes of ethanol. The RNA was finally taken up in double distilled water for experimental use. Northern blot analysis was carried out as described by Church and Gilbert (1984) using Hybond-N (Amersham) nylon membrane and randomly primed probe with a specific activity of approximately 10^9 cpm per ug (Feinberg & Vogelstein, 1984). The *tk* probe was an 840 bp *Pst*I fragment of pTK1 which overlaps 806 bp between the point of fusion with the *Mup* promoter region and a *Pst*I site internal to the TK coding region (Wagner et al., 1981; McKnight et al., 1981). Marker DNA was prepared by run-off transcription of plasmid DNA containing a 1922 bp *Eco*RI-*Pvu*II

fragment of pTK1 inserted into the polylinker region of pT7.1, 10 bp downstream of the T7 RNA polymerase initiation site. The plasmid was digested separately with AccI, BanI, SmaI and NarI and each fragment was transcribed in the presence of low specific activity ^{3}H-UTP to generate fragments 434, 1009, 1306 and 1816 nucleotides long. After recovery these were combined to make a set of markers.

Polyadenylated RNA was separated from the Poly A^- fraction by passing the RNA over an oligo dT column.

Transgenic animals and lines.

(C57BL/6J X CBA) F_1 females were superovulated and DNA microinjected as described in Al Shawi et al., (1988). Transgenic mice were identified by Southern blot analysis of tail DNA. Transgenic females were mated to (C57BL/6J X CBA) F_1 males. Copy number estimates of the transfected plasmid in cell lines were determined by Southern blot analysis.

Primer extension assay.

In a siliconised microcentrifuge tube were mixed 18 fmoles of DNA primer, end labelled with T4 polynucleotide kinase, with 1.8 fmoles of TK poly A^+ RNA, 2ul of 5x Hybridisation buffer (2M NaCl, 50mM PIPES pH 6.4, and water to a final volume of 10ul. The mixture was taken up and sealed in a glass capillary tube. The capillaries were

heated at 70°C for 3 mins and left at the annealing temperature of 60°C for 6 hours. After annealing the contents were expelled into 90 ul of extension reaction buffer (50mM Tris-Hcl pH 7.6, 10mM DTT, 6mM $MgCl_2$, 0.5mM each of dATP, dCTP, dGTP, dTTP, 2.5 ug of actinomycin D and 15 units of AMV reverse transcriptase (Pharmacia FPLC pure). The extension reaction was carried out at 42°C for 60 minutes and terminated by the addition of 5ug of E.Coli tRNA, 0.1 volumes of 3M sodium acetate pH 7.0 and 2.5 volumes of ice cold ethanol. After an overnight precipitation at -20°C the samples were centrifuged at 12000g for 10 minutes, rinsed with 95% ethanol and spun for a further 5 minutes. The pellet was dried under vacuum for 10 minutes and loaded on to a 6% urea acrylamide denaturing gel in 3ul of 1x TBE and 3ul of formamide dye mix.

S1 nuclease protection assay.

In a siliconised centrifuge tube were mixed 20 fmoles of the S1 probe and 2.3 fmoles of TK poly A^+ RNA in 50ul of water. This was precipitated overnight at -20°C by the addition of 0.1 volumes of 3M sodium acetate pH 7.0 and 2.5 volumes of ethanol. After centrifugation at 12000g for 10 minutes the resulting pellet was washed with 70% ethanol, lyophilised, and resuspended in 10ul of 80% formamide, 0.4M NaCl, 1mM EDTA and 50mM PIPES pH6.4. The solution was drawn up into a 20ul siliconised glass capillary, sealed, and immersed into a 75°C water bath for 5 minutes to denature the DNA. These were then transferred immediately to the hybridisation temperature of 60°C and left for 6 hours. The

EX BIBL. UNIV. EDINBURGH

reactions were snap frozen in dry ice and the contents expelled into 190ul of ice cold S1 mix containing 0.28M NaCl, 50mM sodium acetate, 4.5mM ZnSO4, 2ug of denatured salmon sperm DNA and S1 nuclease at 750 units /ml (Boehringer Mannheim). The tubes were incubated at 37^{o}C for 45 minutes and the reaction terminated by the addition of 5ug of E.Coli tRNA, 0.1 volumes of 3M sodium acetate pH7.0 and 2.5 volumes of ethanol. The samples were then treated as described for the primer extension products.

CHAPTER 7: CONSTRUCTION OF HYBRID GENES.

Thymidine kinase constructs

Plasmids are based on psv2gpt (Mulligan & Berg, 1980) retaining the 341 bp PvuII-HindIII SV40 enhancer and early promoter region and the 2293 bp PvuII-EcoRI portion of pBR322, containing the origin of replication and the β-lactamase gene for selection of the recombinant. Plasmids pSVBS2tkSupF and pSVBS2NeoSupF were constructed by myself. In this thesis the letters pSV.. denote an SV40 early promoter and enhancer. To accentuate this the letters EP may be added to give pSVEP. Both nomenclatures refer to the same plasmid.

pSVBS2tkSupF

Between the HindIII site and the EcoRI site the following fragments were inserted (a * denotes that the site was blunt ended with DNA Polymerase I and ligated with T4 DNA Ligase). Seven nucleotides of polylinker M13tg131, HindIII-*BamHI (TGGGATC) (Kieney et al., 1983), a 2283 bp *EcoRI-Sau96I fragment of Mup gene BS2 (Clark et al., 1982; 1985) containing the TATA box and CAP site (see Figures 6 & 7), a ten nucleotide fragment *XbaI-*BamHI from polylinker M13mp19 (CTAGAGGATC) (Yanisch et al., 1985) and a 1759 bp *BglII-*BstEII fragment of Herpes simplex virus thymidine kinase type 1 gene (HSVtk) from plasmid pTK1 (Wilkie et al., 1979). The plasmid pSVBS2tkSupF was completed

with a 10 bp *SalI-*BamHI fragment of the M13tg130 polylinker (TCGACGGATC), a 400 bp AluI-XbaI fragment of PIVX containing the E.Coli Sup-F gene (Seed, 1983) and a 36 bp XbaI-EcoRI fragment of the M13tg130 polylinker (Figures 6 & 8). Restriction analysis of the resultant recombinant plasmid is shown in Figure 9. The major components are delineated. The coordinates of the restriction sites can be compared to the sites used in the cloning strategy of the construct (Figure 8).

pSVBS6tkSupF

Between the HindIII and EcoRI sites of pSV2gpt the following fragments were inserted: an approximately 2200 bp HindIII-Sau3A fragment of Mup gene BS6 (Clark et al., 1982) containing the TATA box and CAP site (see Figures 6 & 7), seven nucleotides of polylinker (GATCCCC),and a 1759 bp *BglII-BstEII fragment of HSVtk from plasmid pTK1 (Wilkie et al., 1979). The plasmid pSVBS6tkSupF was completed with a 10 bp *SalI- *BamHI fragment of PIVX and a 36 bp XbaI-EcoRI fragment of M13tg130 polylinker.

pSVtkSupF and pOtk

Plasmids were constructed from the plasmid pSV2S3BS6tk (Bishop, unpublished).

pSVtksupF retains the pBR322 and SV40 sequences from pSVgpt. The 1759 bp *BglII-*BstEII fragment from pTK1 is linked to the HindIII site of the SV40 enhancer and early promoter by a 20 bp

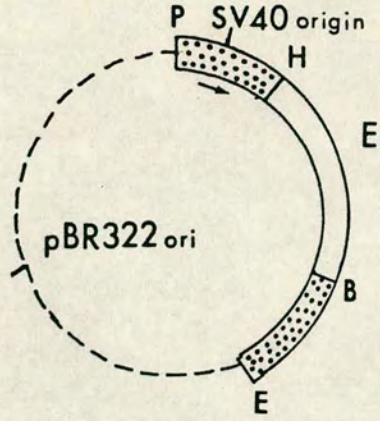
Figure 6.

Structures of pSV2gpt and derivative plasmids: pSVEPtkSupF, pSVBS6tkSupF, pSVBS2tkSupF, and pOtk.

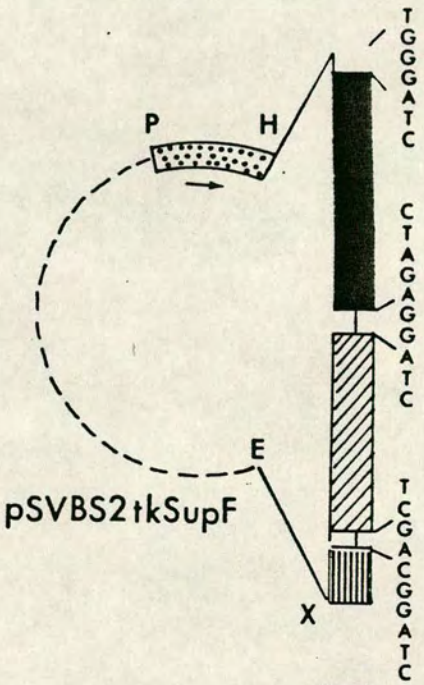
The broken lines indicate 2293 bp of pBR322 vector DNA and the stippled regions represent the 341 bp SV40 enhancer and early promoter region. Major urinary protein (Mup) coding sequences, approximately 2 kb of BS6 or BS2, are shown as shaded boxes. The thymidine kinase (TK) coding region (1759 bp) is shown as a hatched box and the bacterial SupF gene (400 bp) is represented by a cross hatched box. In pSV2gpt the clear region indicates the E.Coli gpt gene. The sequences of the intervening polylinker sequences are shown. The maps are not drawn to scale. The arrow indicates the direction of transcription.

E, EcoR1. H, HindIII. S, Sau3A. P, PvuII. X, XbaI. B, BglII.

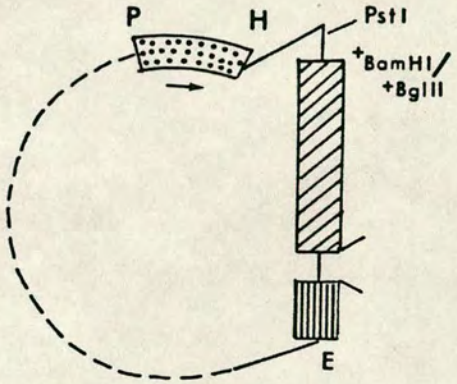
* Indicates that the restriction site was blunt ended with DNA Polymerase 1.



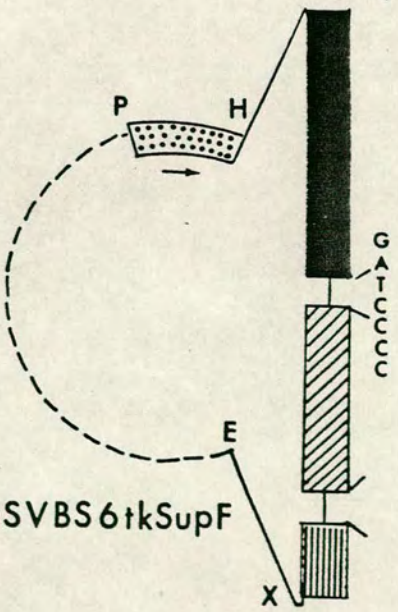
pSV2gpt



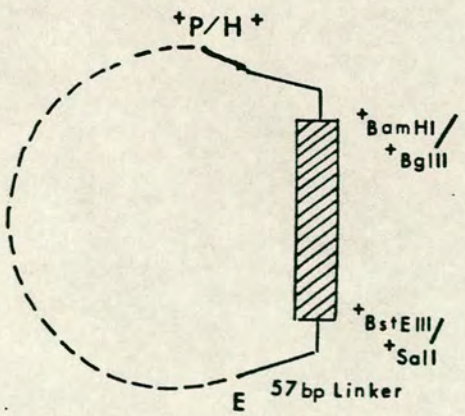
pSVBS2tkSupF



pSVeptkSupF



pSVBS6tkSupF



pOtk

Figure 7.

Schematic representation of constructs pSVBS6tkSupF and pSVBS2tkSupF.

The four major components of the plasmid constructs are shown. These include:

(i) the entire Simian virus 40 (SV40) transcriptional control fragment containing the SV40 enhancer, early promoter and TATA sequence.

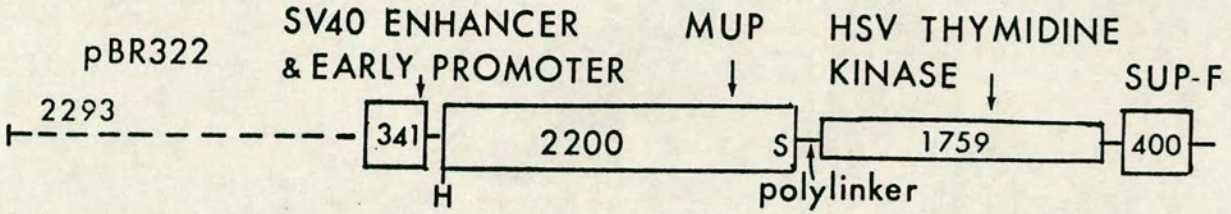
(ii) A fragment of major urinary protein (Mup) gene BS6 or BS2, containing TATA and CAP consensus, with approximately 2 kb of 5' flanking region.

(iii) The Herpes simplex virus thymidine kinase type 1 gene as a reporter.

(iv) The bacterial SupF gene for the isolation of recombinant clones.

The numbers refer to the length in base pairs of each particular segment. The DNA sequence at the Mup/thymidine kinase junctions are shown. The intervening polylinker sequences are underlined (------). The Mup TATA boxes are shown (_) and the putative transcription initiation sites are indicated (arrow).

The sequences shown in the figure for plasmid pSVBS6tkSupF are predicted rather than having being determined from DNA sequencing work that I have undertaken. The portion of the sequence of pSVBS2tkSupF detailed, has been verified by sequencing of the S1 probe which spans this region (See Chapter 9).



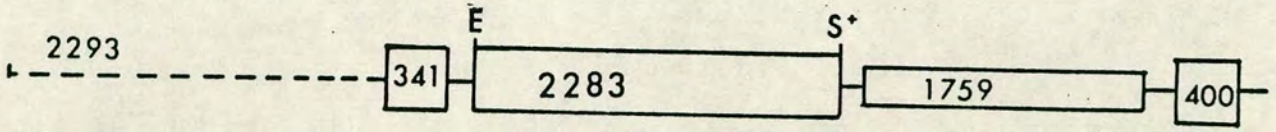
pSVBS6 tk SupF

GAGGAGACCC ATACGGGAAG AGGGAAAAAA AAAAAAACAA AACAAACAAC AACAACAAAA

AAAAAAAACC CGCTGAACCC AGAGAGTATA TAAGGACAAG CAAAGGGGCT GGGGAGTGGA

GTGTAGCCAC GATCCCCGAT CTTGGTGGCG TGAAACTCCC GCACCTCTTT GGCAAGCGCC

TTGTAGAAGC GCGTATGGCT TCGTACCCCT GCCATCAACA CGCGTCTGCG TTCGACCAGG



pSVBS2 tk SupF

AGCCTGACAG AGGTAGAGTC GACCCATACA GGAAGAAAAA AAAAAAAAAA ACCCACTGAA

CCCAGAGAGT ATATGAGGAC AAGCAAAGGA GCTGGGGAGT AGAGTGTAGG CAACATCACC

AGAAAGACGT GGTCCTAGAG GATCGATCTT GGTGGCGTGA AACTCCCGCA CCTCTTTGGC

AAGCGCCTTG TAGAAGCGCG TATGGCTTCG TACCCCTGCC ATCAACACGC GTCTGCGTTC

Figure 8.

Cloning strategy of the plasmid construct pSVBS2tkSupF.

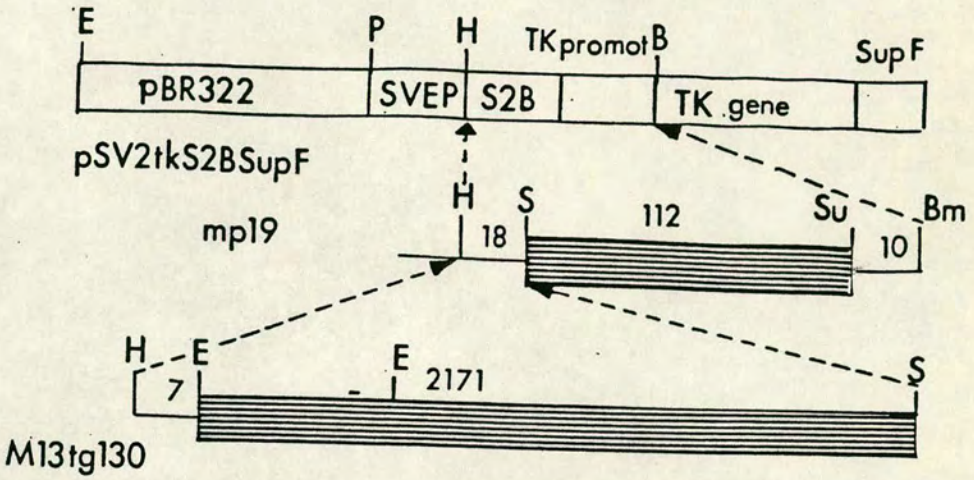
The shaded areas represent the Mup gene BS2 sequences which were cloned in two separate stages into the HindIII and BglII sites of pSV2tkS2BtkSupF. The single lines represent polylinker sequences. The length of each segment in base pairs is indicated. The map is not drawn to scale.

E, EcoRI. P, PvuII. H, HindIII. B, BglII. S, SalI. Su, Sau96. Bm, BamHI.

Figure 9.

Restriction analysis of the plasmid pSVBS2tkSupF.

The diagram shows the coordinates of major restriction enzyme sites which delineate the separate fractions of the construct. The total length of the plasmid is 7146 base pairs.



pBR322 vector	EcoRI	-2
	ScaI	515
	PstI	754
SV40	PvuII	2295
	SphI	2367
	SphI	2439
MUP BS2	HindIII	2635
	EcoRI	2642
	EcoRI	3842
	SalI	4813
linker	Sau96	4925
	Sau3A	4931
	Sau3A	4935
TK gene	MluI	5022
	EcoRV	5225
	EcoRV	5331
	SphI	5386
	SstI	5443
	PstI	5745
	SmaI	6104
	NarI	6614
	BamHI	6706
PIVX SupF	EcoRI	6834
	EcoRI	7037
	BamHI	7051
	PstI	7081
	BglII	7105
	XbaI	7110
pBR322	KpnI	7120
	SphI	7126
	EcoRV	7130
	SstI	7138
	SmaI	7143
	EcoRI	7146

length of M13 polylinker containing the PstI and SalI restriction sites and ending with a *BamHI site (Figure 6). This construct is completed as pSVBS6tkSupF.

pOtk is constructed from the plasmid pSVtk. This is linearized with a PvuII partial digest, cut with HindIII, blunted with DNA polymerase 1 and recircularized to give pOtk. The plasmid does not contain PIVX SupF sequences. It is completed with a 57 bp *SalI-EcoRI fragment from M13 polylinker (Figure 6).

Neomycin Constructs

Plasmids are based on pSV2gpt (Mulligan & Berg, 1980) retaining the 341 bp PvuII-HindIII SV40 enhancer and early promoter region and the 2293 bp PvuII-EcoRI portion of pBR322 containing the origin of replication and the β-lactamase gene.

pSV2tekS2BSupF=B1

Between the EcoRI sites and HindIII sites the following fragments are inserted: twenty one nucleotides of polylinker Mp9 HindIII-Sau3A, a 268 bp Sau3A-BamHI fragment of Mup gene BS6, (Clark et al., 1985), a 458 bp BamHI-BglII fragment of from the plasmid pTK1 (Wilkie et al., 1979) containing the HSVtk promoter, a 1003 bp fragment BglII-SmaI fragment of the neo gene coding for resistance to the antibiotic Neomycin, from plasmid pSV2neo (Southern & Berg, 1982) and a 595 bp fragment SmaI-

*BstEII, from pTK1 containing the HSVtk polyadenylation signals. The plasmid was completed with a 10 bp *SalI-*BamHI fragment of the M13tg130 polylinker, a 400 bp AluI-XbaI fragment of PIVX containing the E.Coli Sup-F gene and a 36 bp XbaI- EcoRI fragment of M13tg130 polylinker.

pSVBS2NeoSupF

This is constructed from B1 by substitution of the 1026 bp HindIII- BglII fragment, containing the Mup BS6 Sau3AI fragment and the HSVtk promoter region, with a 2303 bp HindIII-*BamHI fragment from pSVBS2tkSupF. This contains 10 bp of linker sequence, 2283 bp of Mup BS2 sequence and 10 bp of M13mp19 linker (see Tk constructs).

pSVBS6NeoSupF

This is constructed from B1 as for pSVBS2NeoSupF. Between the HindIII and *BglII sites left after removal of the 1026 bp fragment, the following fragments were inserted: a 1820 bp HindIII-BamHI fragment from BS6-2 and a 323 bp BamHI-Sau3A fragment from BS6-2-4 (Clark et al., 1982). Three base pairs of linker Sau3A-*SmaI complete the sequence.

Construction of the plasmid pSVBS2tkSupF.

In the previous paragraphs I describe the linear order of DNA fragments and sequences which make up each of the DNA constructs. In this section I will describe in detail how the plasmid pSVBS2tkSupF was constructed. This text refers specifically to Figure 8. The starting plasmid was pSV2tkS2BSupF constructed by M. Richardson.

Plasmid pSV2tkS2BSupF contains:

i) a 2293 bp EcoRI-PvuII portion of pBR322.

ii) a 341 bp PvuII-HindIII portion of SV40 DNA containing the SV40 enhancer and early promoter region.

iii) a 1026 bp HindIII-BglII DNA fragment. This fragment contains the MUP BS6 promoter region identified by two Sau3A sites (S2) and extending to a BamHI site (B), 3' to the CAP site (Clark et al., 1985b). See page 55. In addition this fragment contains the HSVtk promoter region delineated by a BamHI-BglII fragment (Wilkie et al., 1979).

iv) a 1759 bp BglII-*BstEII DNA fragment containing the HSVtk gene and v) the bacterial SupF gene (see page 53).

In the construction of pSVBS2tkSupF I have substituted the 1026 bp HindIII-BglII DNA fragment of pSV2tkS2BSupF for approximately 2.2 kb of BS2 sequence containing the BS2 TATA box, potential CAP site and 5' flanking region. In all other respects pSVBS2tkSupF is identical to pSV2tkS2BSupF.

The cloning of the BS2 fragment was achieved in two stages (Figure 8). A * denotes that the site was blunt ended with DNA polymerase I and ligated with T4 DNA ligase.

1) The majority of the BS2 sequence present in a 2171 bp *EcoRI-SalI DNA fragment (Clark et al., 1982) was cloned into M13tg130 by utilising the SalI and *BamHI sites present in the polylinker of the vector. 2) A second 112 bp SalI-*Sau96I DNA fragment contains the BS2 TATA box and potential CAP site. This fragment was cloned into the vector PUC19 (mp19) by utilising the SalI and *XbaI sites present in the polylinker sequence.

The larger BS2 fragment was isolated from M13tg130 in a HindIII-SalI fragment. This was inserted 5' to the BS2 SalI-*Sau96I fragment by utilising the HindIII site present in the PUC19 polylinker (Figure 8). The complete BS2 5' region can then be isolated from PUC19 in a 2300 bp HindIII-*BamHI DNA fragment. This fragment was used to substitute the 1026 bp HindIII- *BglII DNA fragment of pSV2tkS2BSupF. The resultant *BamHI/*BglII fusion at the 3' end of the fragment generates a Sau3A site (Figure 9; coordinate 4935). The sequence of the construct across this junction was verified by sequencing of the S1 probe (See Chapter 9 and Figure 14). The sequence for pSVBS6tkSupF is a predicted sequence and has not been derived from sequencing work that I have undertaken.

CHAPTER 8: TRANSIENT AND STABLE TRANSFECTION ASSAYS.

Introduction.

On initial transfection of a gene into a cell population, only a small proportion of the cells are found to have stably integrated and expressed the gene. The integration of transfected DNA is not essential for its expression. In certain cell lines a relatively high proportion of cells express the transfected gene up to 72 hours post transfection. After this time the activity decreases (Pellicer et al., 1980). This has been termed transient expression. Transient expression systems assay the expression of transfected DNA which has not integrated into the recipient chromosome. They have proved to be an efficient and accurate way to delineate proximal and distal regulatory sequences responsible for correct initiation of RNA transcription (Walker et al., 1983; Charney et al., 1985; Elkareh et al., 1985; Grosschedl & Baltimore, 1985). Since the DNA is not incorporated into the genome the influences of cis-acting sequences at the site of integration are avoided. These may act by silencing or enhancing gene expression. The time of maximum transient expression varies for the type of gene being transfected and the recipient cell line used.

The expression of a selectable transfected gene which has integrated into the recipient genome can be analysed by counting the number of transformants which survive in a selective medium and develop into

colonies. The number of surviving colonies which derive from a population of cells can be expressed as a transformation efficiency.

Banerji et al., (1981) show that the fidelity of transcription in transient assays is as good as that obtained with stable transformants. The performance of transient expression systems can be enhanced by utilising DNA vectors that include portions of the SV40 genome, containing the SV40 replicative origin and early promoter region. These components can facilitate increased expression by replication and consequent copy number amplification. The coupling of an enhancer for example SV40 to a gene can promote transcription (Banerji et al., 1981; DeVilliers & Schaffner, 1981). This depends on the combination of enhancer, promoter and recipient cell.

When introduced into BHKtk^- cells, sequences from the Mup gene BS6 promoter do not initiate transcription unless linked to an SV40 enhancer (Bishop, unpublished). For this reason a SV40 enhancer is also present in the analysis of the BS2 promoter.

The ability of various reconstructed thymidine kinase genes to express RNA and active enzyme was tested using two distinct techniques. Transient levels of tk specific RNA and thymidine kinase enzyme (TK) activity induced in transfected BHKtk^- cells 48 hours post transfection was measured. In addition the ability of recombinant plasmids to transform BHKtk^- cells to a stable TK^+ phenotype in selective HAT growth medium was

determined. In this case the levels of tk mRNA and thymidine kinase enzyme activity in individual clones was measured.

Transient Expression Assay.

Time Course.

Hamster BHKtk$^-$ cells were transfected with the plasmids pSVtkSupF, pSVBS6tkSupF, pSVBS2tkSupF and pOtk. Cultures were harvested at various times post transfection and used to determine TK enzyme levels. The optimum time for expression was found to be 36 hours for the BHKtk$^-$ cells used (Figure 10).

The effects of different Mup gene transcriptional control sequences.

The effects of different exogenous transcriptional control sequences on Herpes simplex virus thymidine kinase gene expression are shown in Table 1. BHKtk$^-$ cells were transfected with the plasmids and TK enzyme activity was measured.

Transient HSV thymidine kinase levels.

Thymidine kinase assays reveal a low level of activity for pSVBS6tkSupF which is significantly different (P 0.055) from the value obtained with the plasmid pSVBS2tkSupF (Table 1B). The amount of enzyme produced in the presence of the BS6 promoter is approximately 20 fold less than that which occurs when transcription

Figure 10.

Transient expression of the HSVtk gene in hamster cells. Time course.

The graph shows the transient HSVTK enzyme activity in lysates of BHKtk$^-$ cells transfected with various recombinant plasmids containing the Herpes simplex virus thymidine kinase type 1 gene. The plasmids pSVEPtkSupF, pSVBS6tkSupF and pSVBS2tkSupF were transfected as described in "Materials & Methods" using 5ug of DNA per $2x10^5$ cells. The DNA-calcium phosphate coprecipitate was removed from the cells after 12 hours and the medium replaced with fresh non-selective medium. The cells were harvested and assayed for HSVTK activity at the time intervals shown.

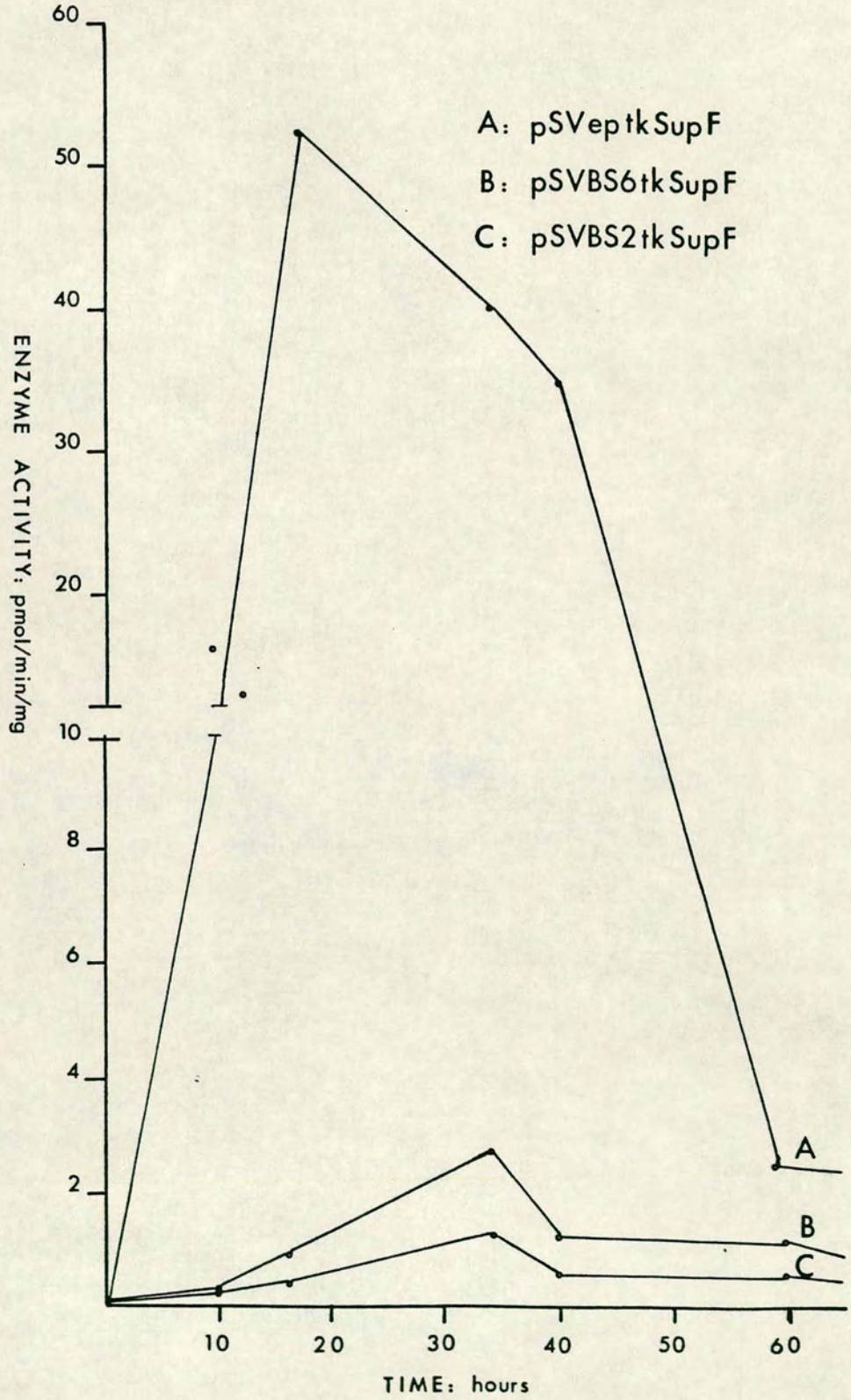
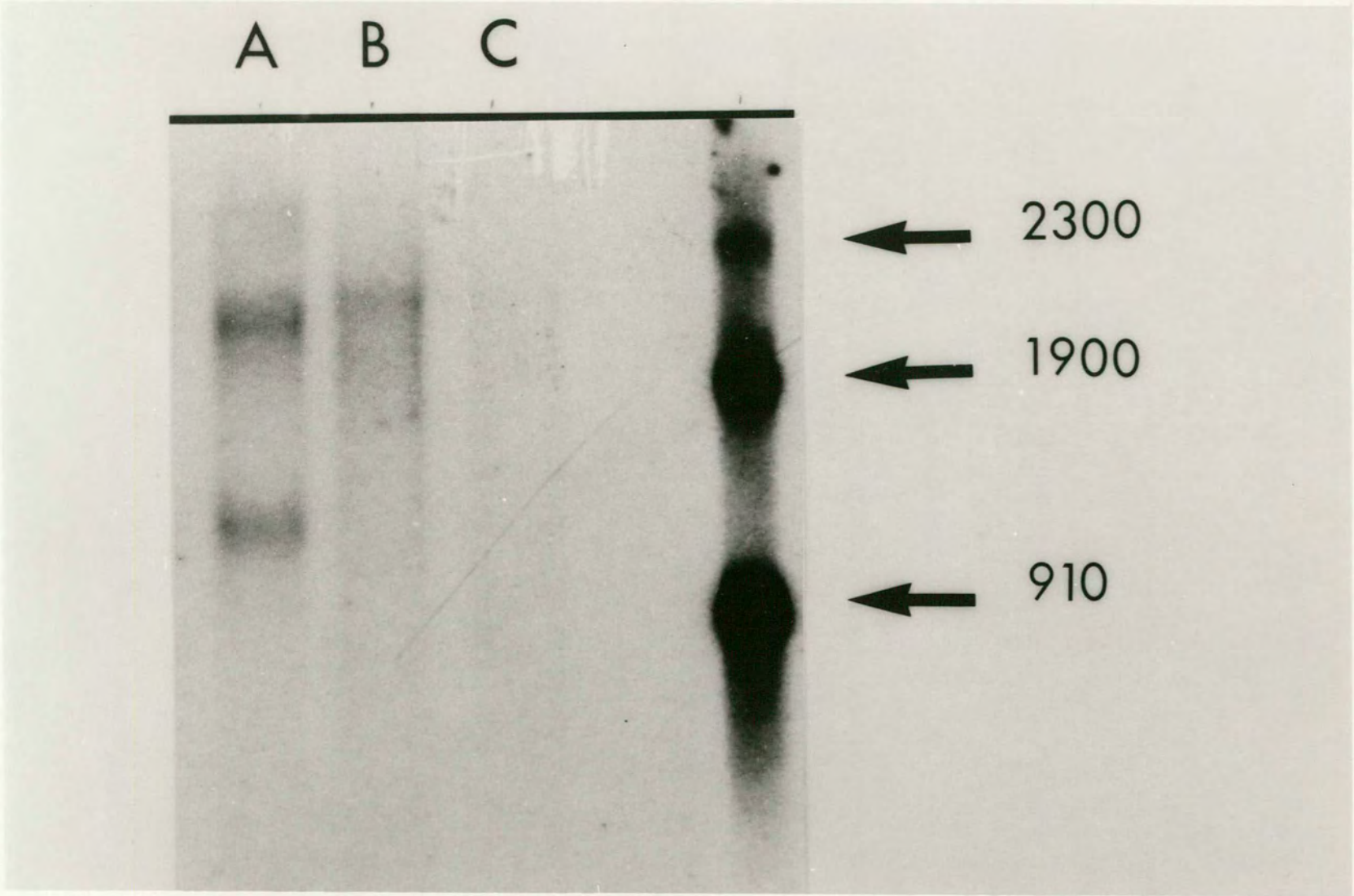

A: pSVeptkSupF
B: pSVBS6tkSupF
C: pSVBS2tkSupF
ENZYME ACTIVITY: pmol/min/mg
60
50
40
30
20
10
8
6
4
2
A
B
C
10
20
30
40
50
60
TIME: hours

Figure 11.

Northern blot analysis of transiently expressed tk specific mRNA.

Thirty-six hrs post transfection, RNA was prepared from cell cultures transfected with the plasmids (A) pSVeptkSupF, (B) pSVBS6tkSupF, (C) pSVBS2tkSupF. 25ug of each was fractionated on a 1.8% agarose-formaldehyde gel, blotted onto a Hybond-N membrane and probed with a 840 bp PstI fragment from pTK1. Marker tracts are transferrin (2300) and MUP (1900 & 910).


A
B
C
2300
1900
910

of the tk sequence is under the control of the SV40 early promoter (Table 1A). The transient activity of the plasmid pSVBS2tkSupF does not differ significantly from the background, pOtk, levels when the t-test for paired comparisons is applied (Table 1B). The low level transient expression of HSVTK from the BS2 promoter could be indicative of a non functional promoter or a very weak level of expression which is undetectable in the assay system.

Transient HSV thymidine kinase levels.

Total mRNA was extracted from duplicate plates and probed with a 840 bp HSVtk specific DNA fragment (PstI-PstI; see Materials & Methods). Figure 11 shows the transient levels of tk specific RNA from cells transfected with the different plasmids. The messenger RNA produced in cells transfected with the plasmid pSVtksupF is approximately 1.3 kb long. This is consistent with transcription initiating at the SV40 early promoter CAP site and terminating at the thymidine kinase polyadenylation signal. No tk mRNA of a size consistent with initiation at the Mup CAP site is detectable in the transient samples for cells transfected with pSVBS6tkSupF or pSVBS2tkSupF. The lack of any transient HSVTK activity or tk-specific RNA in cells transfected with the plasmid pOtk is consistent with the lack of a promoter in this construct.

Taken together the data in Figure 11 and Table 1 show that in these experiments the most sensitive method for measuring promoter activity in transient expression is to quantify the amount of

Table 1A.

A table to show the transient expression of HSV thymidine kinase in BHKtk$^-$ cells transfected with plasmids containing Mup gene transcriptional control elements linked to the HSVtk gene.

Donor DNA (15ug) was added to recipient cells as a calcium phosphate coprecipitate in the absence of carrier as described in Materials & Methods. Fresh medium was added 12 hours later and incubation continued at 37^{o}C, 5% CO_2 for a further 36 hours after which the cells were harvested.

$ HSVTK activity was assayed as described in Materials & Methods. The relative enzyme activities for each construct are shown. HSVTK activity for pSVEPtksupF transfected cells averaged 11pmol/min/mg. All values are relative to this amount.

Table 1B.

A table to show the significance of the differences in transient thymidine kinase values of different constructs in relation to one another.

The method of paired comparisons was used to determine the values of t.

TABLE 1 (A)

Cell Line	£Donor DNA	$Relative thymidine kinase activity over background					
		1	2	3	4	5	Mean
BHK.Tk$^-$	pSVEPtkSupF	100	100	100	100	100	100
BHK.Tk$^-$	pSVBS6tkSupF	2.85	4.9	5.9	6.9	6.6	5.43
BHK.Tk$^-$	pSVBS2tkSupF	1.8	4.5	4.7	2.9	4.4	3.66
BHK.Tk$^-$	pOtk	1.5	2.6	4.5	2.7	4.2	3.1

TABLE 1 (B)

t-test: paired comparisons

Comparison	*t value	[N-1] DF	P
pSVBS6tkSupF/pSVBS2tkSupF	2.8	4	0.055
pSVBS6tkSupF/pOtk	4.49	4	0.015
pSVBS2tkSupF/pOtk	1.67	4	>0.1
pSVEPtkSupF/pOtk	174.9	4	<0.001

*Significance at the 5% level with 4 degrees of freedom requires a t value of 2.776.

thymidine kinase enzyme present.

Transformation of thymidine kinase deficient cells to a TK^+ phenotype.

A second way of measuring gene expression of tk plasmids is to measure the efficiency of biochemical transformation of cells from a TK^- to a TK^+ phenotype. This was carried out by transfecting BHKtk^- cells and counting the number of TK^+ transformed colonies after 14 days growth in selective HAT medium. The effects of different Mup transcriptional control sequences on tk gene activity are represented by a value of transformation efficiency (Table 2). Also shown are the results obtained for the transformation of BHKtk^- cells to a neomycin resistant phenotype. The constructs contained the same Mup control elements as did the tk plasmids, but differed in having the bacterial neo gene as a reporter. To allow for the variation in transformation efficiency, the results are expressed as a percentage of the number of colonies obtained with pSVtksupF for the tk constructs and plasmid B1 for the neo series.

The results from the transformation assay are different from those observed for the transient expression assay.

Expression of tk in cells transfected with pSVBS2tkSupF is shown by the survival of cells in HAT. The transformation efficiency of pSVBS2tkSupF is 16% relative to pSVtkSupF. The value obtained with

Table 2.

The number of colonies appearing in selective medium after the transfection of BHKtk$^-$ cells with various constructs containing MUP gene transcriptional control sequences linked to the HSVtk gene or the bacterial neomycin gene.

Donor DNA (5ug) was added to 2 x 10^5 cells as a calcium phosphate coprecipitate in the absence of carrier as described in the Materials & Methods.

* The numbers of colonies represent the total colonies obtained with four duplicate transfections for each construct. Thymidine kinase expressing colonies were selected in 10% HAT. Neomycin selection was carried out with the antibiotic G418 (400ug/ml).

$ Relative efficiencies are expressed as a percentage of the number of colonies obtained with plasmid pSVtksupF for the tk series and B1 for the neo constructs.

TABLE 2

Cell type	#Donor DNA	*Colony formation per experiment			Total	$Relative Efficiency (%)
		1	2	3		
BHk tk$^-$	None	0	0	0	0	0
BHk tk$^-$	pSVEPtkSupF	100	300	150	550	100
BHk tk$^-$	pSVBS6tkSupF	80	94	105	182	33
BHk tk$^-$	pSVBS2tkSupF	26	51	11	88	16
BHk tk$^-$	pOtk	2	6	0	8	1.45
BHk tk$^-$	None	0	0		0	0
BHk tk$^-$	pSV2tekS2BSupF (B1)	396	237		633	100
BHk tk$^-$	pSVBS6neoSupF	297	93		390	67
BHk tk$^-$	pSVBS2neoSupF	162	34		196	31

pSVBS6tkSupF is 33% (Table 2). The BS6 sequences are twice as efficient as the BS2 sequences in transforming cells to a TK^+ phenotype. When the neo gene was used as a reporter the same relationship between BS6 and BS2 sequences emerged. However, for a particular promoter sequence, the neo series of plasmids do tend to have a greater transformation efficiency (approximately 50%) than those harbouring a tk reporter sequence for example compare pSVBS6tkSupF and pSVBS6NeoSupF (Table 2). This observed difference may be due to the ability of relative promoter strengths to be influenced by sequences in the genes transcribed from them. This observation has previously been described when different heterologous sequences are attached to the HSVtk promoter (Novak & Rothenberg, 1986; Koltunow et al., 1987). However, as the ratio of BS6 versus BS2 transformants are the same for each reporter sequence this is unlikely to be the case. Alternatively lower levels of neo expression may be enough to save cells from G418 compared with the tk levels needed to save the cells from HAT.

There is a low but real transformation of cells from TK^- to a TK^+ phenotype when transfected with the plasmid pOtk. As this plasmid lacks any natural promoter the observed activity is probably a result of the activation of the tk sequences by promoter elements at the site of integration.

Gene expression in stably transformed cells.

In order to examine the fidelity of transcription from the Mup BS6

and BS2 CAP sites several independent colonies from the transformation experiments were cloned and propagated in HAT medium. Cultures were harvested 14 days after isolation and used to determine levels of tk specific RNA and TK enzyme activity (Table 3). The thymidine kinase probe will hybridise proportionately to the same amount of tk RNA in the total cellular RNA population as is present in the same intensity band of the marker. The tkRNA in a population of 25ug total cellular RNA gives the same signal as 0.12ng of tkRNA synthesised from a 1816 bp EcoR1-Nar1 fragment placed downstream of a T7 promoter. For example in clone HS9 the amount of tk messenger RNA constitutes around 0.0005% of the total mRNA.

The HSVTK activity in different clones is the same.

Surprisingly, despite the large differences between different plasmids observed in the transient assay, the values of HSV thymidine kinase activity in stably transfected clones of pSVtkSupF, pSVBS6tkSupF, pSVBS2tkSupF and pOtk are approximately the same. The amount of tk specific RNA does however vary widely between clones (Figures 12 & 13). The average value is 100 pmol/min/mg (Table 3). Thus, although markedly different levels of tk gene expression are measured by transient and transformation assay, long term culture in selective HAT medium may select cells in which the level of tk gene expression is further controlled to a suitable level.

Table 3.

Thymidine kinase activities of individual clones transfected with different recombinant plasmids and the size of the tk specific mRNA produced from each clone.

* The sizes of the transcripts were measured against thymidine kinase DNA transcribed from the T7 RNA polymerase initiation site. A discrepancy of 1mm in the measurement of the bands leads to an error of 50 base pairs in the size obtained

HSVTK activity was assayed as described in Materials & Methods.

TABLE 3

Donor DNA	Name of Clone	HSV Thymidine kinase Activity (pmol/min/mg)	*Size of RNA transcript (base pairs)
pSVEPtkSupF	SV 2	52.8	1580
	SV 3	109	1550
	SV 4	98	1520
pSVBS6tkSupF	HS 1	120	1470
	HS 2	86	1470 + 1180
	HS 5	250	1470
	HS 6	116	1470 + 930
	HS 9	ND	1470
pSVBS2tkSupF	ES 1	188	ND
	ES 2	162	1390
	ES 3	118	1500
	ES 6	193	1700 + 1250
	ES 8	ND	1260
	ES 9	147	<1816
	ES11	81	1340
pOtk	Ot 1	202	1650
	Ot 3	165	1550
	Ot 6	163	1580

Figure 12.

Northern blot analysis of stably transfected clones.

The figure shows a Northern blot hybridisation of tk specific RNA from clones stably transfected with plasmid constructs pSVBS2tkSupF, denoted clones ES2, ES3 etc, and construct pSVBS6tkSupF, denoted HS1, HS2 etc. 25 ug of total cellular RNA from each transformed cell line grown under selective conditions was fractionated on a 1.8% agarose formaldehyde gel, blotted onto Hybond-N membranes and probed with a 840 bp Pst1 fragment from pTK1. The marker tracts are the products of run-off transcription of a 1922 bp EcoRI-PvuII fragment from pTK1 inserted downstream from the T7 RNA polymerase initiation site. The plasmid was digested separately with AccI, BanI, SmaI and NarI and each fragment transcribed to give the fragment lengths indicated.

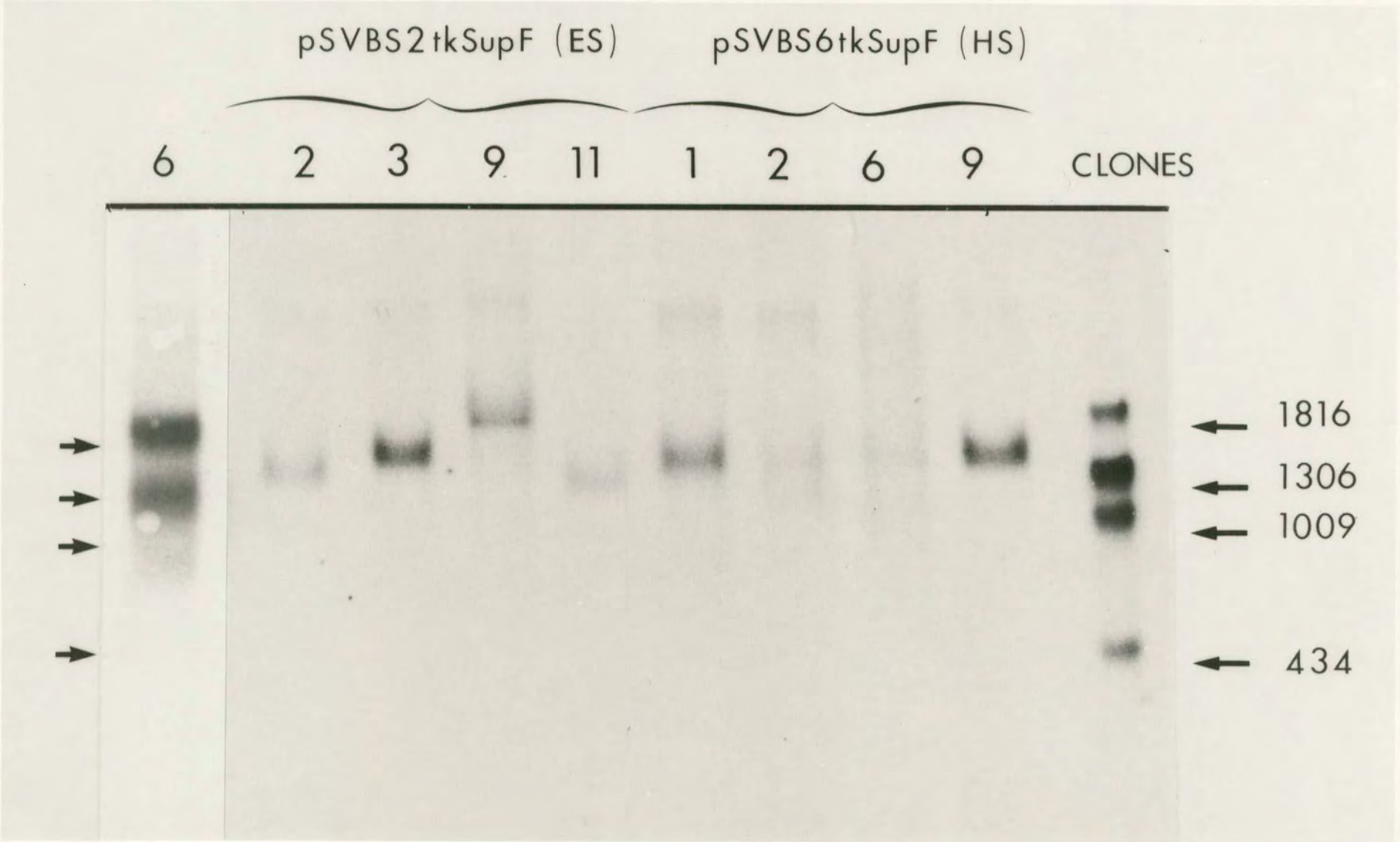
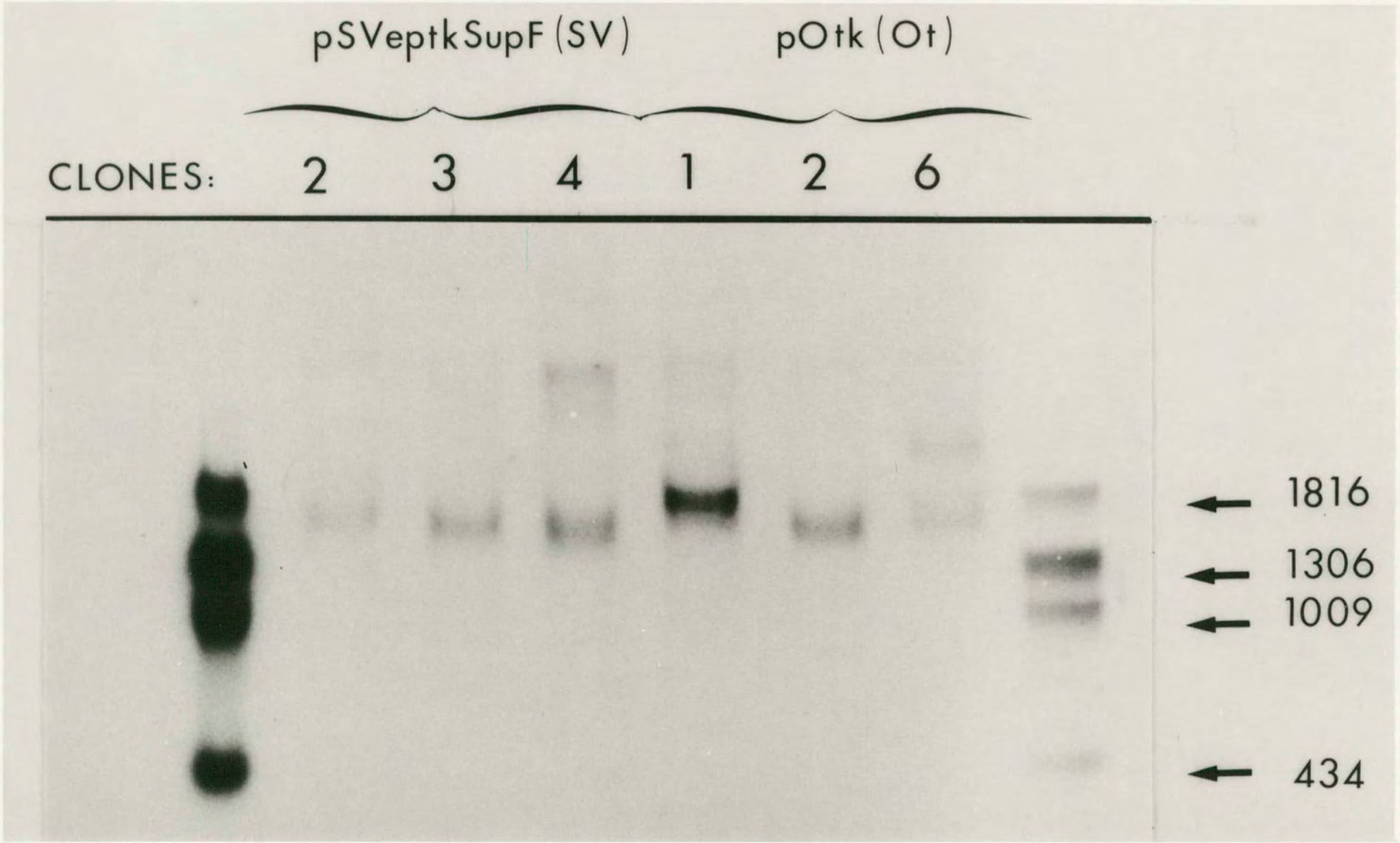

pSVBS2tkSupF (ES)
pSVBS6tkSupF (HS)
6
2
3
9
11
1
2
6
9
CLONES
1816
1306
1009
434

Figure 13.

Northern blot analysis of stably transfected clones.

The figure shows a Northern blot hybridisation of tk specific RNA from clones stably transfected with plasmid constructs pSVtkSupF, denoted clones SV2, SV3 etc, and construct pOtk, denoted Ot1, Ot2 etc. 25 ug of total cellular RNA from each transformed cell line grown under selective conditions was fractionated on a 1.8% agarose formaldehyde gel, blotted onto Hybond-N membranes and probed with a 840 bp Pst1 fragment from pTK1. The marker tracts are the products of run-off transcription of a 1922 bp EcoRI-PvuII fragment from pTK1 inserted downstream from the T7 RNA polymerase initiation site. The plasmid was digested separately with AccI, BanI, SmaI and NarI and each fragment transcribed to give the fragment lengths indicated.


pSVeptkSupF (SV)
pOtk (Ot)
CLONES: 2 3 4 1 2 6
1816
1306
1009
434

RNA transcripts in clones transfected with pSVBS6tkSupF and pSVBS2tkSupF.

Figures 12 & 13 show the sizes of tk-specific mRNA in clones of cells transfected with the different Mup/tk hydrid plasmids. The clones are titled ES if transfected with the plasmid pSVBS2tkSupF, and HS if transfected with the plasmid pSVBS6tkSupF. The letters ES denote the 2283 bp DNA fragment EcoRI-Sau96I from Mup gene BS2. The letters HS denote the 2200 bp HindIII-Sau3A DNA fragment from Mup gene BS6 (Figure 8). Clones transfected with the plasmids pSVtkSupF and pOtk are named SV and Ot respectively. Table 3 gives the sizes of these transcripts as measured against RNA markers produced by run-on transcription of tk DNA fragments from the T7 RNA polymerase initiation site (Figure 14). An HSV-specific tk mRNA of approximately 1.5 kb was present at roughly the same concentration in a number of different clones transfected with pSVtkSupF (Figure 13). This strongly suggests that the tk gene is being transcribed from the SV40 promoter in the transformed cell lines SV2, SV3 and SV4, since the normal tk mRNA is known to have a size of 1.3 kb and initiation from the SV40 CAP site and termination at the normal tk polyadenylation site would result in an RNA of similar size. In cells transfected with pSVBS6tkSupF a band of 1.47 kb is present in all of the clones. In addition clones HS2 and HS6 express extra faint trancripts of length 1180 bp and 930 bp respectively. These shorter messages could be due to initiation of tk mRNA from cryptic transcription start sites within the tk coding region or they may be initiating from the reverse strand of

tk. The observed 1.47 kb transcript is consistent with the initiation of the tk mRNA from the BS6 CAP site and termination at the tk polyadenylation signals. These results confirm previous studies with the plasmid pSVBS6tkSupF (Ghazal, 1986). S1 mapping of the 5' ends of transcripts synthesised from the pSVBS6tkSupF construct transiently expressed in BHK cells, show that the SV40 early promoter is used in pSVtksupF while the Mup promoter is preferentially used in pSVBS6tkSupF. The difference in size between the predicted message from the BS6 and SV40 early CAP sites and the observed lengths presumably reflects heterogeneity at the 3' end of the RNA.

The sizes of transcripts in clones stably transformed with pOtk and pSVBS2tkSupF are very heterogeneous.

There is considerable heterogeneity in the sizes of transcripts from individual clones obtained after transfection with pSVBS2tk SupF (Figure 12). The sizes range from 1.3 kb in clone ES11 to more than 1.8 kb in clone ES9. Clone ES6 has two messages of 1700 and 1250 bp (Figure 12). None of these transcripts is consistent with the initiation of transcription at the BS2 CAP site and termination at the tk polyadenylation sequences. This would generate a 1.45 kb transcript. A similar variation in transcript size is observed in the stable transformants of pOtk (Figure 13). The size of these transcripts all exceed 1.4 kb.

Large, faintly hybridising tk specific transcripts are seen in transfected clones.

Northern blot analysis of RNA reveals several large RNA species which hybridize weakly to the tk probe in some of the clones (Figure 12 & 13). In the cases of plasmids pSVBS6tkSupF and pSVBS2tkSupF these transcripts could be initiating from the SV40 early promoter located 2 kb upstream. Probing of the Northerns with a 340 bp PvuII- HindIII fragment specific for the SV40 promoter and with the whole plasmid, minus the tk sequence, do not reveal cross hybridisation with any of these larger transcripts (data not shown). The extra long transcripts must therefore be transcribed from the chromosome or the reverse tk strand. The presence of the large transcripts in clones transfected with pSVtkSupF also suggest the latter case. Probing of the filters with a coding strand specific ribo-probe would be useful in resolving this problem.

CONCLUSIONS.

Promoter strength and transformation efficiency.

It has been suggested that the initial, transient rate of expression of a gene is a major determining factor in the subsequent rate of biochemical transformation (Gorman et al., 1983; Lang et al., 1983; Spandidos & Wilkie, 1983; Sodroski et al., 1984; Rutherford & Nienhuis, 1987). Thus, the proportion of cells which express the tk gene above the threshold for survival in HAT should depend on

the relative expression of the gene and this is reflected in the levels of transiently expressed enzyme. This may not always be the case as the influence of sequences at the site of integration can overcome the lack of expression of a transfected gene for example pOtk.

Promoter preference in plasmids transfected with pSVBS6tkSupF.

The initiation of transcription at the CAP site of BS6 in preference to the SV40 early promoter is unusual. Wasylyk et al., (1983) and de Villiers et al., (1982) show that in plasmids containing an SV40 enhancer and more than one natural or substitute promoter, proximal promoters are activated in preference to distal ones. Studies by Kadesh & Berg, (1986) lead to the same conclusion although another effect termed promoter occlusion may distort the magnitude of the apparent preference. Promoter occlusion causes reduced transcription at a downstream promoter if transcription is initiated at a nearby upstream promoter.

The proximity of the SV40 promoter to the SV40 enhancer does not make it the preferential transcription initiation point. The sizes of the tk specific mRNAs produced in stably transfected clones of pSVBS6tkSupF are consistent with initiation of transcription in the region of the distal Mup promoter sequences, as shown by Ghazal (1986). This is in contrast to the results described above. It has been suggested that the Mup sequences may in some way down regulate the SV40 promoter (Ghazal, 1986). This may result from the

binding of a *trans*-acting factor to 5' *Mup* sequences which prevent the entry of DNA polymerase II to the template at the SV40 CAP site.

The *Mup* result may not be unique. Lang *et al.*, (1983) showed that when an LTR enhancer is placed upstream of *tk* under the control of its own promoter, insertion of the LTR promoter between the enhancer and *tk* promoter had no effect in modulating gene expression. These results suggest that there is no absolute preference for the proximal promoter. Alternatively this observation might relate specifically to the *tk* sequence.

CHAPTER 9: S1 NUCLEASE PROTECTION AND PRIMER EXTENSION ASSAYS.

The Northern blot data shows that the tk specific RNAs from the clones stably transfected with the plasmid pSVBS2tkSupF are very heterogeneous. They are not of a size consistent with initiation from the BS2 CAP site and termination at the tk polyadenylation sequences. In order to map the transcription initiation start sites, S1 nuclease and primer extension assays were implemented.

Results.

S1 nuclease protection experiments.

The S1 probe was prepared by cloning the 1542 bp EcoRI-SphI fragment from pSVBS2tkSupF into the EcoRI and SphI sites of the M13tg131 polylinker. Synthesis of the second strand, complementary to the RNA, was initiated from the M13 universal priming site located 3' to the SphI site. The strand was uniformly labelled by the incorporation of alpha-dCTP P^{32} in the synthesis reaction. The resulting double stranded DNA was restricted with HpaI and EcoRV and the resulting 1015 bp fragment isolated from a 0.8% agarose gel. Hybridisation of the probe with the RNA was effected in 80% formamide at 60^{o}C (see Materials & Methods). The probe is 59.6% GC rich. The 434 bp RNA transcribed from the EcoRI-AccI fragment of pTK1 placed 10 bp downstream of the T7 transcription initiation site, served as an S1 control. The observed

290 bp protected fragment in the control reaction was in agreement with the expected result (see Figure 14 & 15 and Table 4A). The single stranded S1 probe is not informative for transcripts which may be transcribed from the antisense strand of the tk gene.

Poly A^+ RNA from each of the four clones stably transfected with pSVBS2tkSupF were hybridised to the S1 probe and exposed to S1 nuclease. In each case the size of the protected fragment was different, as measured against an M13tg131 sequencing ladder (Figure 15). RNA from clone ES2 resulted in a protected fragment of 124 bp (Table 4A). The doublet bands observed in clones ES3 (280 + 276 bp) and ES9 (293 + 291) are probably due to attack by S1 nuclease on the DNA/RNA hybrid. Alternatively they might represent two transcripts which are very similar in size and derive from different copies of the integrated plasmid. Two widely separated fragments of 318 and 150 bp are seen with RNA from clone ES6. This is in agreement with the fact that two major transcipts are seen on the Northern blot (Figure 12). The 5' termini of the RNA transcripts mapped in this way are shown in Figure 17.

Primer extension experiments.

To determine the start sites of transcription in these clones, primer extension assays were carried out. Two synthetic oligonucleotides were used. Primer A is 23 bp long and is 59.6% GC rich. It is complementary to RNA transcribed from the tk coding region between coordinates 5020 and 5042 (see Figure 14a). Primer B

Figure 14.

(A) Restriction map and coordinates (bp) of the plasmid pSVBS2tkSupF showing its relationship to the probes used for S1 protection and primer extension assays. The S1 probe is a 1015 bp Hpa1- EcoRV DNA fragment. In the primer extension assays a 22-mer oligonucleotide (coordinates 5020-5042) or a 20-mer oligonucleotide (5151- 5169) was used. The thymidine kinase coding region of the construct is shown by the cross hatched box.

(B) Restriction map of the 1992 bp EcoRI- PvuII from pTK1 inserted into the polylinker region of pT71, 10 bp downstream of the T7 RNA polymerase initiation site. The plasmid was digested separately with Acc1, Ban1, Sma1 and Nar1 to generate the fragment lengths shown. RNA transcribed from the 434 bp species was used in the S1 and primer extension assays as a control.

(C) Schematic enlargement of the restriction map of the 434 bp Acc1- EcoR1 fragment from pTK1 inserted 10 bp downstream of the T7 polymerase initiation site. The figure shows the extent of hybridisation of the S1 probe to the RNA generated from this fragment. A 290 bp protected fragment is expected. The position of the primer extension probes (PE) are also shown. The 256 and 383 bp species represent the primer extension controls from primer A and B respectively. The lower coordinates represent those in the original plasmid (pSVBS2tkSupF) shown in (A). This map is not to scale.

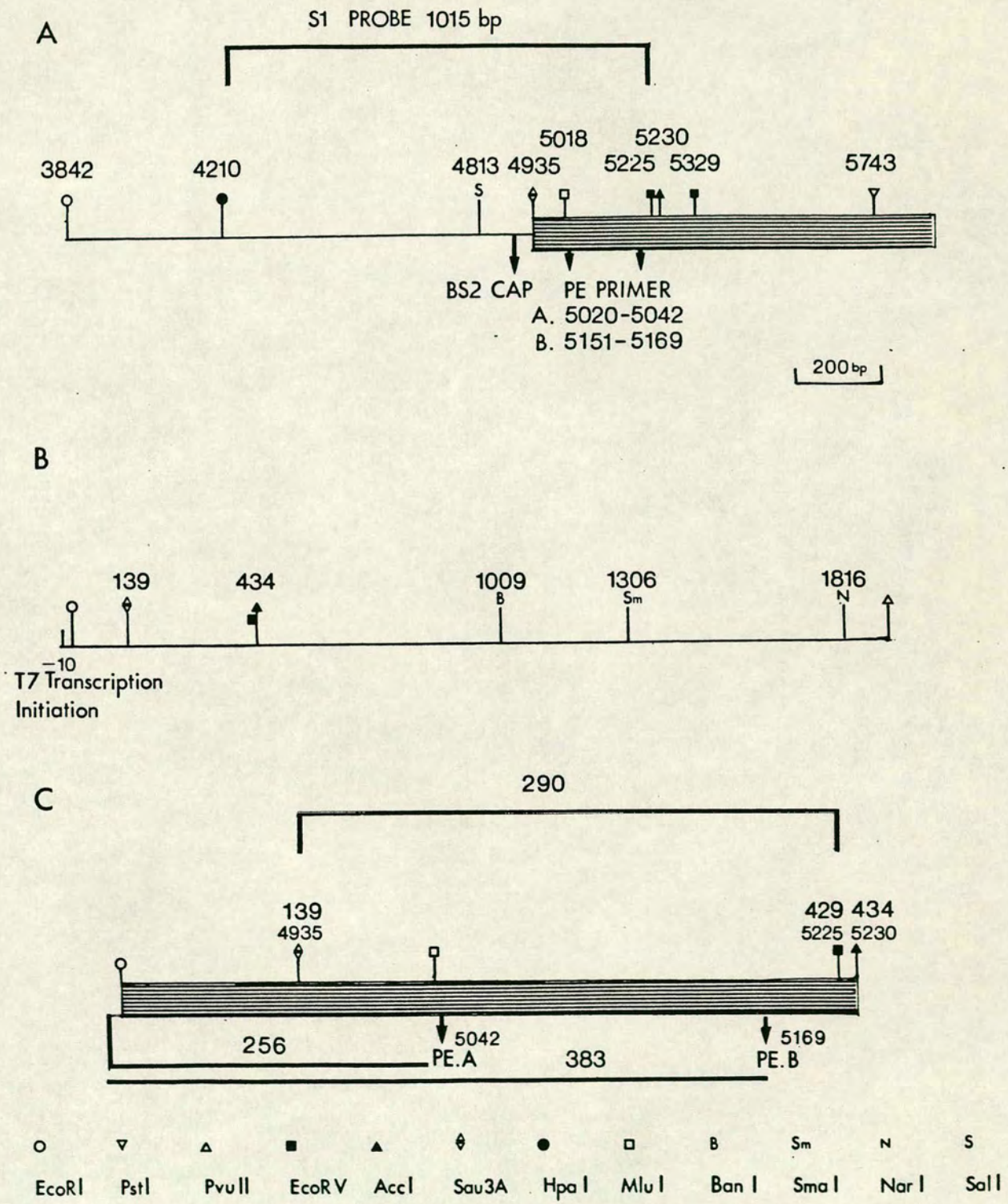

A
S1 PROBE 1015 bp
3842
4210
4813
S
4935
5018
5225
5230
5329
5743
BS2 CAP
PE PRIMER
A. 5020-5042
B. 5151-5169
200 bp
B
139
434
1009
B
1306
Sm
1816
N
−10
T7 Transcription
Initiation
C
290
139
4935
429 434
5225 5230
256
5042
PE.A
383
5169
PE.B
EcoRI
PstI
PvuII
EcoRV
AccI
Sau3A
HpaI
MluI
BanI
SmaI
NarI
SalI

Figure 15.

Electrophoretic analysis of the products of S1 nuclease protection.

Lanes A, C, G, T are a sequence ladder of M13tg131 used to provide molecular weight markers.

Lane 1: S1 protection of yeast RNA (S1 protection control).

Lane 2: S1 probe alone. Hpa1-EcoRV 1015 bp.

Lanes 3 & 4: S1 protection of the RNA from the 434 bp EcoR1-Acc1 fragment of pTK1.

Lane 5: S1 protection of poly A^+ RNA from clone ES2.

Lane 6: S1 protection of poly A^+ RNA from clone ES3.

Lane 7: S1 protection of poly A^+ RNA from clone ES6.

Lane 8: S1 protection of poly A^+ RNA from clone ES9.

Table 4A.

S1 protection of the RNA from the 434 bp EcoRI-HpaI fragment from pTK1 results in a protected band of 290 bp (see Figure 14). Protection of the poly A^+ RNA from clones stably transfected with the plasmid pSVBS2tkSupF yield the fragment sizes indicated.

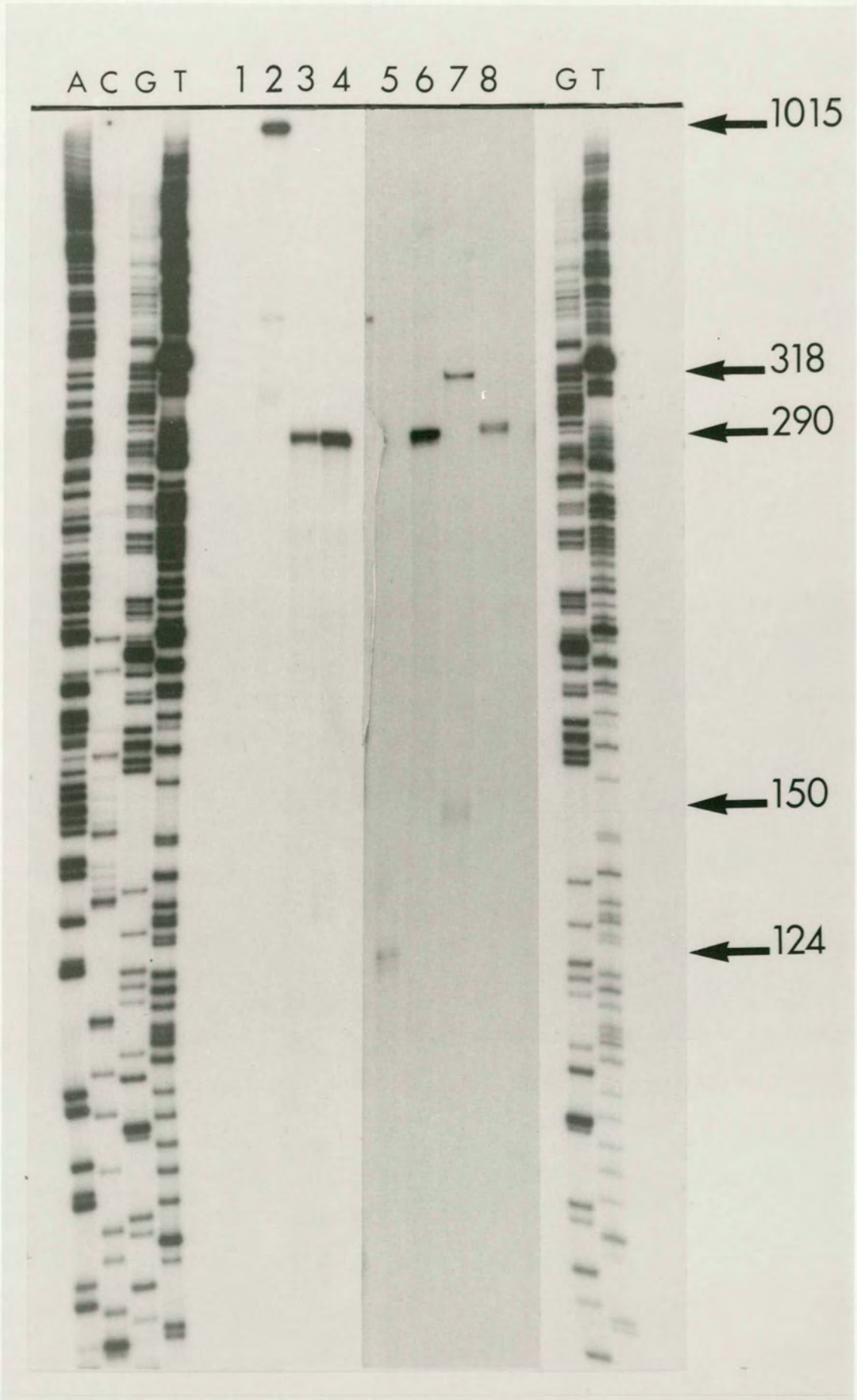

A C G T
1 2 3 4 5 6 7 8
G T
1015
318
290
150
124

Figure 16.

Electrophoretic analysis of the products of primer extension assay.

Lanes A, C, G, T are a sequence ladder of M13tg131 used to provide molecular weight markers.

Lane 1: Primer extension control. Extension product of the 434 bp EcoRI-AccI fragment of pTK1.

Lane 2: Extension product of yeast RNA.

Lane 3: Extension product of poly A^+ RNA from clone ES3.

Lane 4: Extension product of poly A^+ RNA from clone ES6.

Lane 5: Extension product of poly A^+ RNA from clone ES9.

Lane 6: Extension product from the 434 bp EcoRI-AccI fragment from pTK1 with the primer B (5151- 5169).

Lane 7: Primer extension product of $polyA^+$ RNA from clone ES6.

Lane 8: Primer extension product of $polyA^+$ RNA from clone ES2.

Each RNA species was extended with the primer A (5020- 5042), in lanes 1-5 and primer B (5151-5169) in lanes 6-8.

Table 4B.

The table shows the expected sizes of extension products from primer A from extrapolation from the S1 data. Also shown are the sizes of the observed extension products.

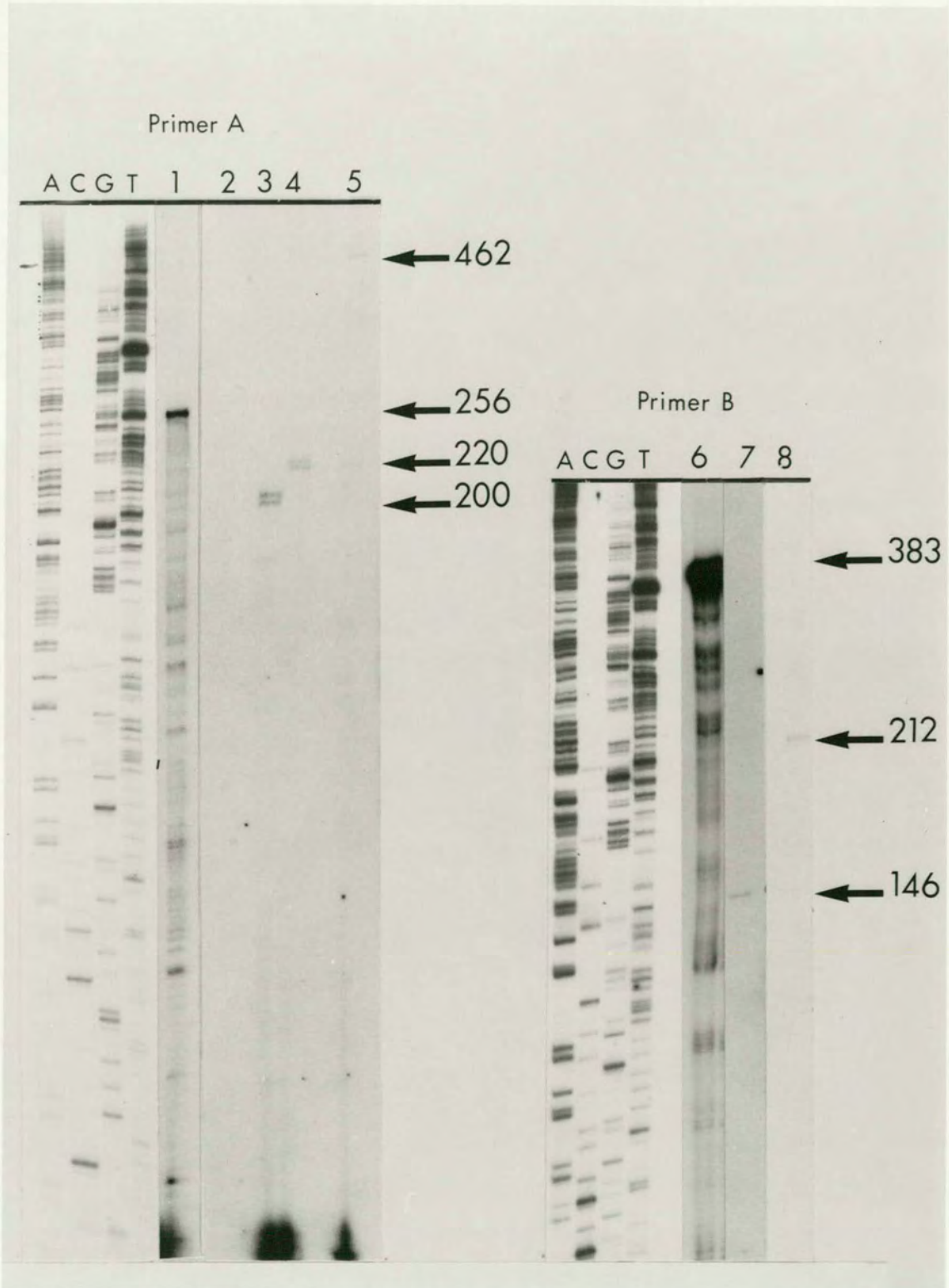

Primer A
A C G T 1 2 3 4 5
462
256
220
200
Primer B
A C G T 6 7 8
383
212
146

Table 4A.

LANE	RNA SPECIES	SIZE OF PROTECTED FRAGMENT(s) (bp)
3 & 4	SI CONTROL	290
5	ES CLONE 2	124
6	ES CLONE 3	280 + 276
7	ES CLONE 6	318 + 150
8	ES CLONE 9	293 + 291

Table 4B.

		SIZE OF EXTENSION PRODUCT(s) (bp)		
LANE	RNA SPECIES	EXPECTED (from SI assay)	OBSERVED PRIMER A (5020-5042)	OBSERVED PRIMER B (5151-5169)
1 & 6	PRIMER EXTENSION CONTROL	256 + 383	256	383
8	ES CLONE 2	*	*	212
3	ES CLONE 3	97 + 93	200 + 196	*
4 & 7	ES CLONE 6	135 + *	220/214 + *	* + 146
5	ES CLONE 9	110 + 108	462	*

is 19 nucleotides long, is 50% GC rich and maps between coordinates 5151 and 5169. The primers were end-labelled to high specific activity using T4 kinase. The annealing reaction was carried out at 60°C (see Materials & Methods). As a control in these experiments the primers were hybridised to the RNA transcribed from a 434 bp EcoRI-AccI fragment of pTK1 inserted downstream of the T7 promoter. The extension of this RNA with Primer A yields a product which is 256 bp long. Primer B yields a 383 bp extension product (Figure 14b & c). These observed results are in agreement with the sizes expected (Table 4B). The products of the primer extension reaction are shown in Figure 16.

The primer extension data does not agree with the S1 data.

Table 4B shows the expected lengths of the extension products obtained with RNA from each of the stably transfected clones. These values were extrapolated from the S1 data. The observed results do not conform with the values predicted from the nuclease protection assays. Each primer extension product is longer than expected. The primer extension results map the start of transcription to sites upstream of those indicated in the S1 assay. This was not an artefact of the reactions as the control RNA in both experiments yielded the expected products of 256 and 383 bp. An explanation is that rearrangement of the construct has taken place. The primer extension data indicates the position of the start of transcription. This data is consistent with the Northern blot data. The S1 data shows the extent of hybridisation of the RNA to a Mup/tk DNA

specific sequence (Figure 17). The difference between the two termini represents sequences present in the transcript which are not a simple continuation of the foreign gene sequence, in the 5' direction, from the point at which homology with the S1 probe ends. In principle this transition could be due to any discontinuity in the input sequence. There are several explanations for this. For example an insertion of DNA into, or deletion of DNA from, the transfected plasmid sequence. Alternatively the transition could represent the change from foreign DNA to chromosomal DNA as a result of the integration of a rearranged plasmid.

In clones ES2 and ES6, interruption of the plasmid sequence has occurred 3' to the point at which Primer A hybridises. Thus the sequences complementary to the primer are not transcribed. This explains why Primer A was uninformative for analysing these clones (Table 4B). However, the break points do occur 5' to the site at which Primer B hybridises. This primer is informative for mapping the transcription start sites in these clones. The distances of the start sites map from 50 bp upstream (clone ES6) to 350 bp upstream (clone ES9) (Figure 18).

None of the tk specific transcripts initiate from the probable BS2 CAP site.

Only one of the tk-specific RNA's present in clone ES6 contains sequences which are complementary to Mup DNA. This is shown by the extent of hybridisation of the RNA to the S1 probe. The Mup region

Figure 17.

A figure to show the sequences at the site of rearrangement in clones ES2, ES3, ES6 and ES9 as mapped by S1 nuclease assay. The arrows indicate the 5' boundary of the RNA/DNA hybrid. Also shown are the sequence and position of the two primers used in the primer extension assay, the thymidine kinase ATG, the Mup TATA box region and the proposed BS2 CAP site. The numbers represent the map coordinates in the original plasmid pSVEPBS2tkSupF.

■ EcoRV ▲ AccI Sau96I Sau3AI

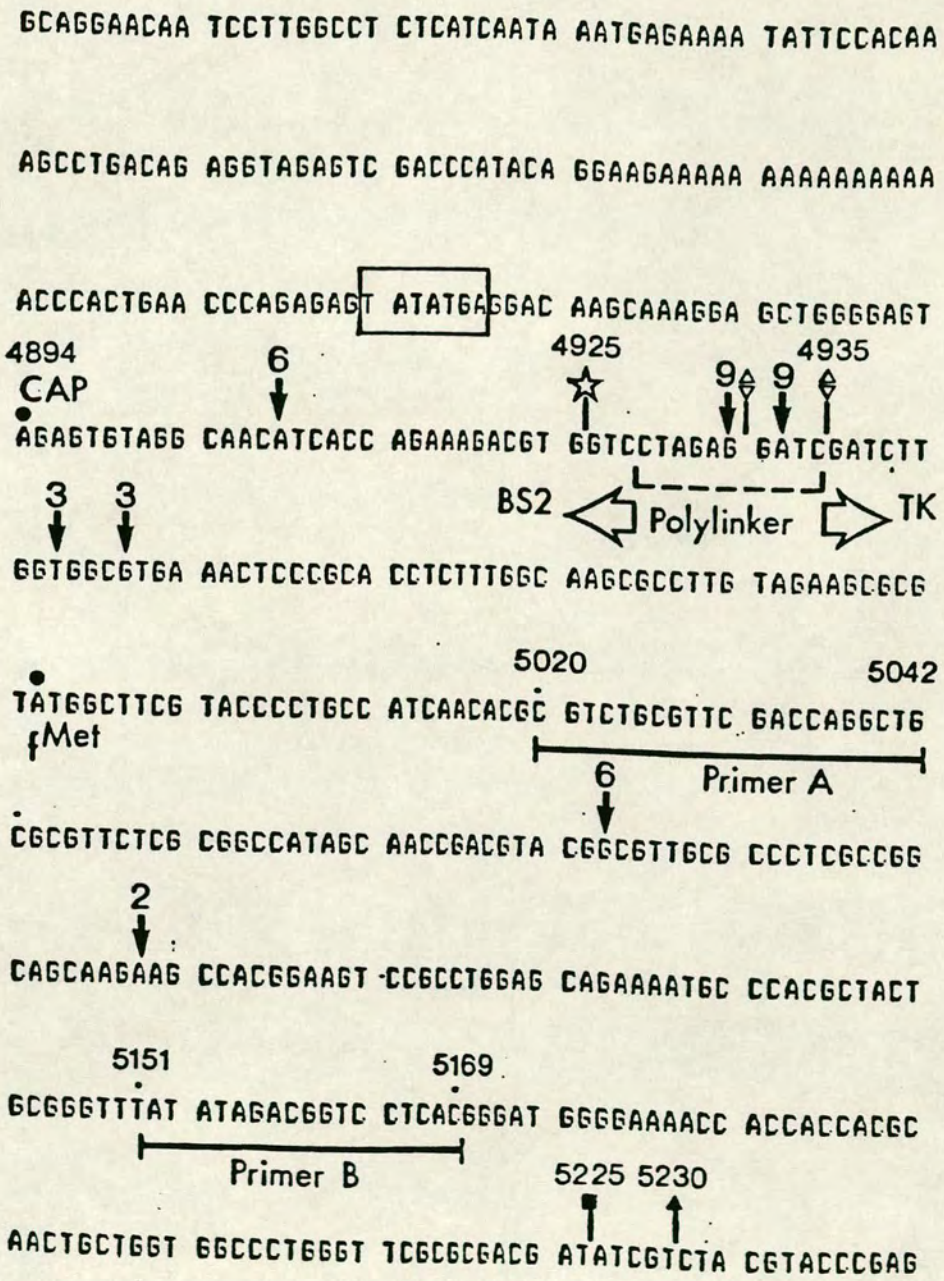
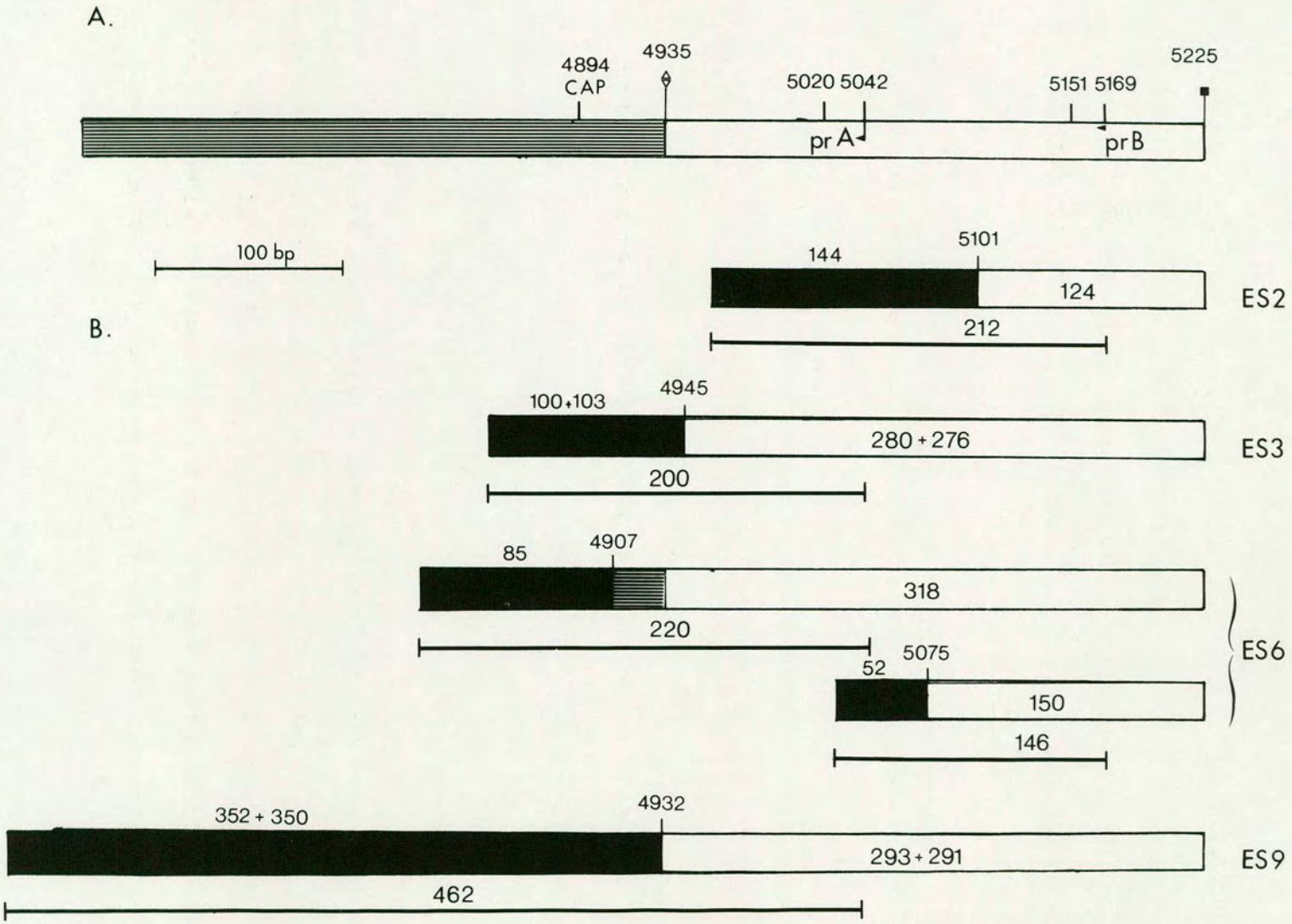

GCAGGAACAA TCCTTGGCCT CTCATCAATA AATGAGAAAA TATTCCACAA
AGCCTGACAG AGGTAGAGTC GACCCATACA GGAAGAAAAA AAAAAAAAAA
ACCCACTGAA CCCAGAGAGT ATATGAGGAC AAGCAAAGGA GCTGGGGAGT
4894
CAP
6
4925
9
9
4935
AGAGTGTAGG CAACATCACC AGAAAGACGT GGTCCTAGAG GATCGATCTT
3
3
BS2
Polylinker
TK
GGTGGCGTGA AACTCCCGCA CCTCTTTGGC AAGCGCCTTG TAGAAGCGCG
5020
5042
TATGGCTTCG TACCCCTGCC ATCAACACGC GTCTGCGTTC GACCAGGCTG
fMet
6
Primer A
CGCGTTCTCG CGGCCATAGC AACCGACGTA CGGCGTTGCG CCCTCGCCGG
2
CAGCAAGAAG CCACGGAAGT CCGCCTGGAG CAGAAAATGC CCACGCTACT
5151
5169
GCGGGTTTAT ATAGACGGTC CTCACGGGAT GGGGAAAACC ACCACCACGC
Primer B
5225 5230
AACTGCTGGT GGCCCTGGGT TCGCGCGACG ATATCGTCTA CGTACCCGAG

Figure 18..

The composite DNA sequences that lead to active thymidine kinase expression in clones of BHKtk$^-$ cells stably transfected with the plasmid pSVBS2tkSupF.

A) The boxed region represents the HSV thymidine kinase coding region separated from the Mup sequences (cross hatched box) by a Sau3AI site. The coordinates of the two primer extension primers A and B are shown.

B) Components of the sequences which give rise to active thymidine kinase expression in the stably transfected clones ES2, 3, 6, 9. The boxed/hatched areas show the length of the protected fragment in base pairs as determined by S1 nuclease mapping. This can incorporate thymidine kinase and Mup sequences. The S1 probe is the 1015 bp HpaI-EcoRV fragment (Figure 14A). The coordinates represent the points of rearrangements of the plasmid in each of the clones (see Figure 17). The true start of transcription is determined by primer extension assay. The amount of "adopted" DNA is given above each shaded box. The length of the extension product is shown underneath each map.


A.
4894
CAP
4935
5020 5042
5151 5169
5225
pr A
pr B
100 bp
B.
144
5101
124
ES2
212
100+103
4945
280 + 276
ES3
200
85
4907
318
220
ES6
52
5075
150
146
352 + 350
4932
293 + 291
ES9
462

which is transcribed constitutes 25 bp (Figures 17 & 18). This region does not include the Mup TATA box or the probable CAP site. RNA from clone ES9 hybridises to all of the tk DNA sequences. The break point in the plasmid DNA is mapped to the polylinker separating the tk and Mup sequences. Similarly RNA from clone ES3 is complementary to all but 13 bp of the tk sequence (Figure 17).

This data suggests that the integration of an intact copy of the plasmid pSVBS2tkSupF does not lead to correct thymidine kinase expression.

The structure of integrated plasmid copies.

The presence of plasmid rearrangements in each of the clones was analysed by enzymatic digestion of cellular genomic DNA. Southern blot analysis of these samples showed the presence of intact and rearranged copies in each cell line. The copy number of the plasmid was determined and this was compared to the level of HSVTK measured.

Many rearrangements are observed in the stably transfected clones.

The structures of the tk integrants in the genomes of stably transfected clones were investigated. The DNA was restricted with HindIII + BamHI or with MluI + BamHI. The plasmid has unique HindIII and MluI sites and two BamHI sites (Figure 19).

Hybridisation of a tk-specific probe to a Southern transfer of MluI and BamHI digested DNA revealed a prominent 1684 bp band in each of the clones (Figure 20). This fragment contains the tk coding regions which are necessary for survival under HAT selection. When the DNA was digested with HindIII and BamHI, a band of 4071 bp which is common to all of the clones was seen. This represents the intact Mup and tk coding sequences. The simplest pattern was observed in a MluI + BamHI digestion of DNA from clone ES9. Only two bands were observed. The predominant 1684 bp band could indicate tandemly arranged copies of the plasmid but the data is only informative for the regions between the two enzyme sites. In all of the remaining clones a large number of strongly hybridising bands are seen (Figure 20). All four clones contain some integral Mup and tk sequences but many rearrangements are observed. Primer extension and S1 analysis have shown that those copies which harbour intact Mup coding sequences are inactive. Of the rearrangements probably only one needs to be actively expressed to ensure survival of the cell under selective pressure. I have probably therefore selected cells containing a productively rearranged plasmid which expresses thymidine kinase, amongst a host of silent intact and nonproductively rearranged copies.

The number of integrated plasmid copies and HSVTK activity.

The copy number of the integrated tk gene in the four stably transfected clones was determined by densitometric scanning of the autoradiogram (Figure 20). Each clone contained on average 6 copies

of the plasmid pSVBS2tkSupF. The highest number (10) was found in clone ES6 and the lowest (4) in clones ES2 and ES9. Clone ES3 contains 8 copies. The copy number of the plasmid was compared directly to the level of HSV thymidine kinase present (Table 5). There is no correlation between copy number and HSVTK activity. Each clone has approximately the same amount of enzyme activity; 150 units/mg protein. Survival of the cell in HAT medium necessitates the expression of only one of the plasmid copies. The constant level of TK activity in all of the clones could suggest that the relative strength of the promoter governing tk expression is about the same. However, the intensity of the bands which represent the products of the primer extension assay are different for each clone (Figure 16). The amount of polyA$^+$ RNA used was the same in each case. The intensity of each band is in direct proportion to the amount of transcript which can be primed in the polyA$^+$ RNA population. Clone ES3 has approximately twice the amount of primable message as ES6 and up to four times the amount of ES9 (Figure 16) yet all the clones have the same TK activity. This could suggest that the translational efficiency of each of the tk specific transcripts is different.

Figure 19.

The figure shows restriction maps of the construct pSVBS2tkSupF showing the sites used to establish the copy number of the plasmid in stably transformed clones. The sizes of the expected fragments are shown.

A). BamHI + MluI digest.

B). BamHI + HindIII digest.

● BamHI ○ EcoRI □ HindIII M MluI ▽ PstI

The underlined MluI-PstI region represents the extent of hybridisation of the 840 bp PstI-PstI fragment probe from PTKI.

C). Schematic diagram of the pSVBS2tkSupF construct showing the delineation of the plasmid components by restriction sites.

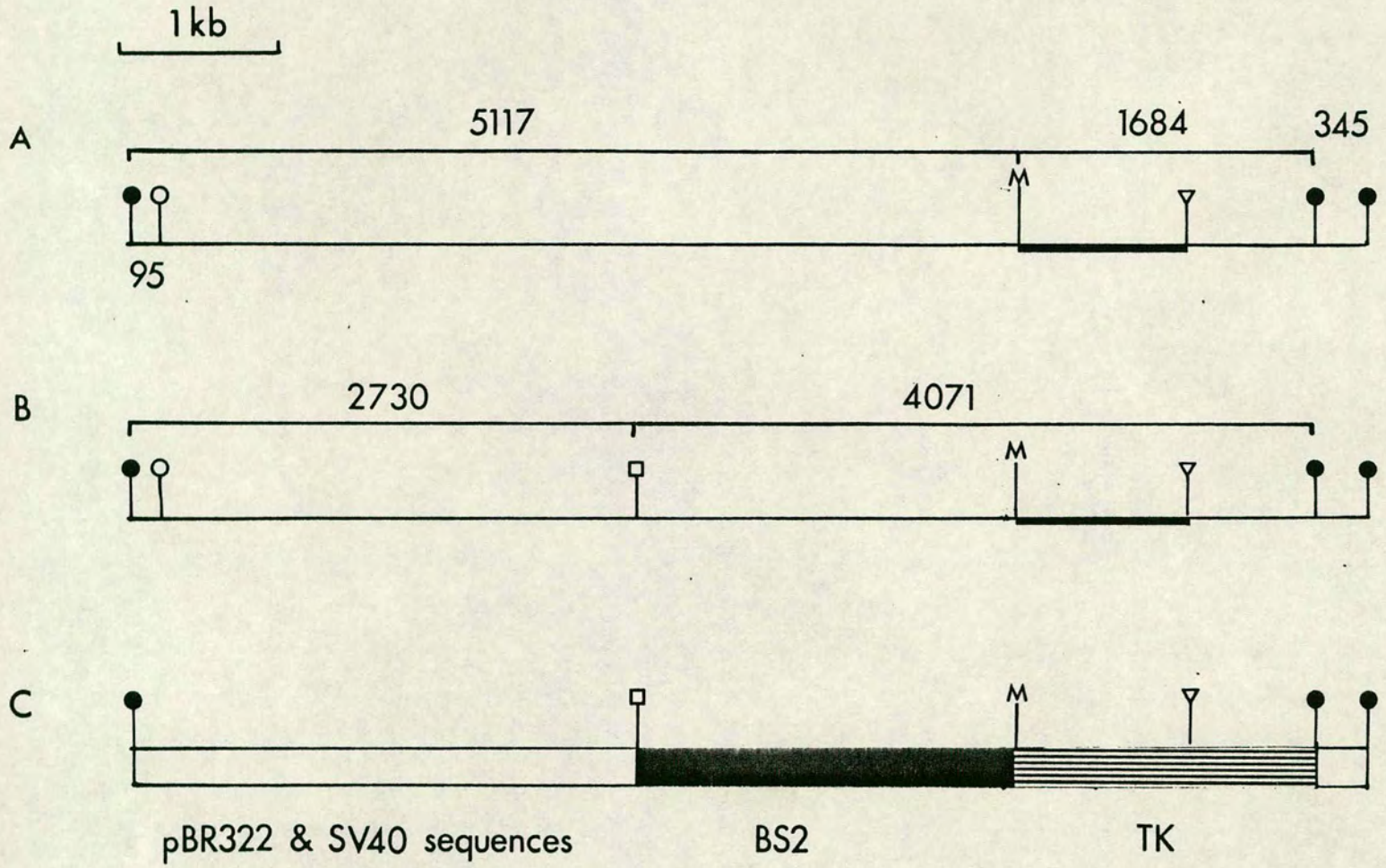

1kb
A
5117
1684
345
M
95
B
2730
4071
M
C
M
pBR322 & SV40 sequences
BS2
TK

Figure 20.

Southern blot of the genomic DNA from each of four clones and transfected with the plasmid pSVBS2tkSupF, restricted with the pairs of enzymes indicated and probed with the 840 bp PstI-PstI fragment from pTK1.

Clone ES9 (9), ES6 (6), ES3 (3) and ES2 (2).

Lane A: 10 copy number marker. The DNA is restricted with HindIII and BamHI. This gives a fragment of 4071 bp and a partial fragment of 4416 bp.
Lane B: 5 copy number marker. The DNA is restricted with MluI and BamHI. This gives a 1684 bp fragment and a 2029 bp partial.
Lane C: 1 copy number marker, as lane B.

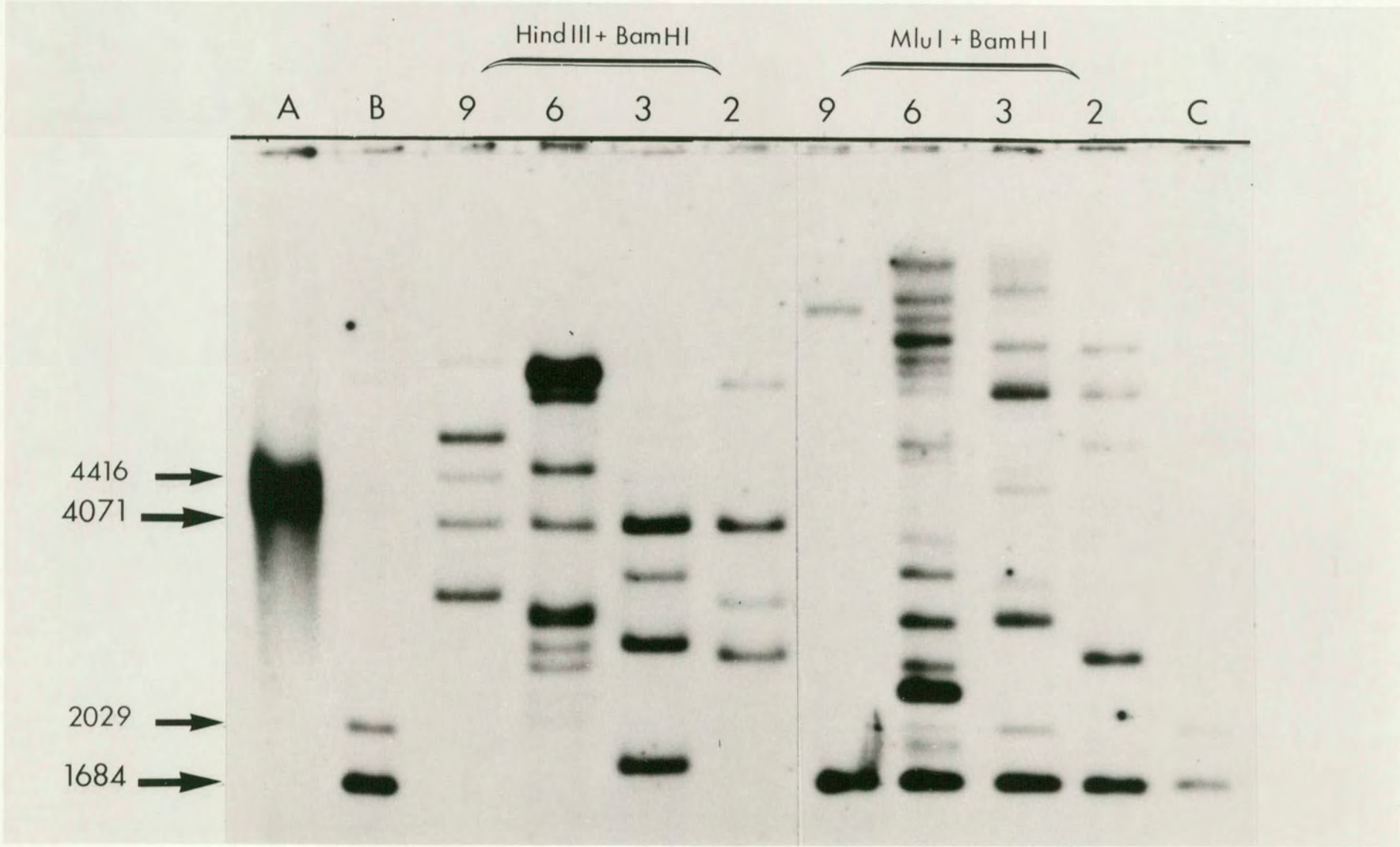

Hind III + BamH I
Mlu I + BamH I
A
B
9
6
3
2
9
6
3
2
C
4416
4071
2029
1684

Table 5.

Clone	Copy Number	HSV Thymidine Kinase activity (pmol/min/mg)	mRNA levels (relative)
ES2	4	162	?
ES3	8	118	4
ES6	10	193	2
ES9	3	147	(1)

Relationship between the copy number of thymidine kinase sequences in cells stably transfected with the plasmid pSVBS2tkSupF and the resultant HSV thymidine kinase activity and relative mRNA levels.

CHAPTER 10.

THE EFFECTIVENESS OF THE Mup BS2 PROMOTER IN THE INITIATION OF TRANSCRIPTION:

DISCUSSION OF TISSUE CULTURE EXPERIMENTS.

The fidelity of expression of transfected genes.

The intrinsic ability of promoter sequences to initiate transcription can be assessed by their fusion to a selectable gene marker and transfection into a recipient cell line. One such selectable gene is the Herpes simplex virus type I gene coding for thymidine kinase. This enzyme is involved in the pyrimidine salvage pathway (Figure 5). Transfection and expression of this gene results in the stable conversion of BHK*tk*$^-$ cells to a TK^+ phenotype. Transfected cells grow in selective HAT medium because of the stable integration of one or more copies of the foreign DNA into the recipient genome. In assessing the function of a promoter by analysing the selectable gene product, the assumption is made that the expression of the gene is usually controlled by its own promoter. However, the effect of the chromosomal environment at the site of integration cannot be discounted. Studies on a number of selectable and non-selectable genes show that the fidelity of gene expression is sufficient to produce properly processed mRNAs of correct size which are polyadenylated and transported to the cytoplasm. However, the analysis of the 5' termini of certain

transfected genes has revealed some anomalies. Wold et al., (1979) studied rabbit β-globin transcripts in transfected L cells. SI mapping revealed a 5' terminus about 45 nucleotides downstream from the natural 5' end. Dierks et al., (1981) also observed these aberrant transcripts but found that the great majority of RNAs were correctly initiated. There is still some question as to whether the truncated transcripts are a consequence of aberrant initiation or of post-transcriptional cleavage of correctly initiated messages. A low level of aberrant 5' termini have also been seen for HSVtk mRNA (McKnight et al., 1981). In this case incorrect initiation was indicated by transcripts which were longer than normal.

Aberrant transcripts following transfection of promoterless tk genes.

Transfection of promoterless tk genes can give rise to HAT resistant colonies (Roberts & Axel, 1982; Pulm & Knippers, 1985). When promoterless tk genes are used in transfection experiments the number of HAT resistant clones is 0.1-1% of the number obtained with an intact HSVtk gene (McKnight & Gavis, 1980; Pulm & Knippers, 1985). Roberts & Axel, (1982) transformed aprt$^-$ tk$^-$ L-cells with a plasmid containing a wild type aprt gene and a truncated, promoterless tk gene. Transformants which integrated a single copy of the plasmid exhibited the aprt$^+$ phenotype but remained tk$^-$ in selective medium (Wigler et al., 1979). An RNA transcript of 0.9 kb was expressed. This is 400 bp smaller than the normal tk message which is 1.3 kb long. Some

transformants containing a single copy of the wild type tk gene also expressed this 0.9 kb message. A third less abundant 1.1 kb message is also expressed by the promoterless plasmid. TK^+ variants which contained a high copy number of the promoterless tk gene were also isolated. These copies have been shown to be organised in tandem arrays. Roberts & Axel, (1982) hypothesise that the truncated 0.9 kb mRNA message is initiated from an internal transcription start site within the tk coding region. Translation of this transcript produces a protein with reduced catalytic activity. Only when it is abundant is sufficient enzyme activity accumulated to permit phenotypic conversion from TK^- to TK^+. Hiscott et al., (1980) demonstrated that amplification was responsible for selection at restrictive temperatures in a cell line containing an SV40 temperature sensitive A gene. Pulm & Knippers (1985) also observed high copy number arrays, 10-25 per genome, of promoterless tk genes. Analysis of the $polyA^+$ RNA from these clones revealed that the major RNA message was of the same size as that from a clone containing the intact tk gene. A smaller aberrant RNA was also detectable but this was not in excess of the longer mRNA species. In addition the thymidine kinase activities in these cell lines were the same as in lines transfected with normal genes. A cell line containing only 2-3 copies of the promoterless tk plasmid gave an RNA message of 1.3 kb. It was concluded that survival of truncated tk genes under selective conditions does not depend on the overproduction of aberrant tk transcripts. Instead at least one of the truncated genes is in optimal association with an active cellular promoter. This provides an

alternative way in which a promoterless tk gene is able to transform a TK^- cell population to TK^+ phenotype. A consequence of this will be that the length of the tk message will depend upon the distance of the promoter from the tk coding sequences.

Hiller et al., (1988) proved that the hypothesis of Pulm & Knippers (1985) was correct. From the low copy number cell line isolated by Pulm & Knippers, they determined the structure of a surrogate promoter by isolating the integrant gene from a genomic library. Nucleotide sequencing data suggested that the transfected circularly closed plasmid DNA was apparently cut at a site immediately upstream of the tk sequence and joined via blunt end ligation to a cellular DNA segment in a reaction involving little if any sequence homology. The cellular DNA flanking the integrated tk sequence functioned as a surrogate promoter, since the transformation efficiency of a plasmid containing the tk sequence with associated mouse DNA was at least as high as the transformation efficiency of a plasmid carrying the intact tk gene with the authentic HSV promoter. Sequencing of the 5' promoter region revealed a TATA box element, two CCAAT boxes and a GC element with no similarity to any of the known thymidine kinase promoters. The CAP site of the cellular promoter has the same spatial arrangement to the coding sequences as the CAP site of the normal intact tk gene. This results in the observed 1.3 kb transcript.

In conclusion, the appearance of HAT resistant colonies in cells

transfected with truncated promoterless genes should indicate the frequency with which active transcriptional promoters have been moved 5' to the integrated gene. In view of the low transformation efficiency of truncated tk genes this type of activation must be a rare event.

Non homologous recombination in mammalian cells:

The site of integration of transfected DNA.

Direct evidence for the integration of the DNA complex into mammalian cell DNA was obtained by Robins et al., (1981). They introduced the human growth hormone gene into rodent cells and showed by in-situ hybridisation methods that the complex is integrated at a specific site on one of the cell chromosomes. In each line the hybridisation was restricted to a single random chromosomal site which was different in different lines. The distribution of integration sites is unlikely to be tru ly random because the input DNA probably does not have equal access to all parts of the genome. The overrepresentation of repetitive sequences at integration sites argues for some unevenness in random integration (Wallenberg et al., 1984 & 1987; Kato et al., 1986). Kinnaird & Bishop (unpublished results) analysed several cell lines stably transformed to a TK^+ phenotype with a BS6/HSV thymidine kinase gene. One of these lines on further examination was found to have integrated the transfected DNA into a middle repetitive sequence of the hamster genome.

Folger et al., (1982) and Brinster et al., (1985) suggest that exogenous DNA may integrate at pre-existing chromosomal breaks which determine and limit the number of integration sites. It may be a feature of repetitive DNA that it is inherently less stable than active gene sequences. This would account for the observation above.

Random integration requires little or no sequence homology at junctions between foreign DNA and chromosomes.

When foreign DNA integrates into a chromosome, it generates a rather distinct junction, usually with an abrupt boundary between foreign and chromosomal DNA. Recently information has become available on the nature of the DNA sequences at the site of integration (Kato et al., 1986). Transformants were obtained by the transfection of mouse tk^- L cells with a plasmid containing the entire thymidine kinase gene together with pBR322 sequences. DNA sequences containing cellular and exogenous DNA junctions in different rescued plasmids were recovered from several of the resultant tk^+ cell lines by molecular cloning. The plasmids were analysed in detail by nucleotide sequencing and hybridisation techniques. No direct or indirect repeats, or any detectable sequence homologies between the exogenous and cellular DNAs near the junctions, were found at the site of integration. The absence of significant homology indicates that random integration probably arises from non-homologous recombination.

The structure of integrated DNA.

The commonest outcome of DNA that is microinjected into mouse eggs is the insertion into one of the chromosomes of a single array of direct tandem repeats (head to tail configuration; see Palmiter & Brinster, 1986). Head to head and tail to tail arrangements, and also rearrangements of the foreign DNA, are rarely observed. Fragments of chromosomal DNA have been found embedded in arrays (Covarrubias et al., 1986; 1987) and insertion is sometimes accompanied by rearrangement in the flanking DNA sequences (Robins et al., 1981; Stringer, 1982; Wilkie & Palmiter, 1987; Mahon et al., 1988). Low concentrations of DNA tend to give transformants carrying fewer copies of the foreign DNA, which are often integrated as single copies at independent sites (Perucho et al., 1980; Robins et al., 1981).

The mechanisms involved in random integration.

The events that accompany DNA integration are not fully understood. However, the observed structures of arrays are consistent with models based on the fate of DNA microinjected of transfected in cell culture. Molecules can become joined in two ways. The ends of linear molecules are joined randomly, by a DNA ligation mechanism, and it is probably in this way that foreign DNA fragments are incorporated into the chromosomal DNA. The structures of integration sites at which linear molecules have joined to chromosomes suggest that DNA free ends stimulate random integration directly. Most commonly,

linear molecules are joined to chromosomal sequences at or near the original ends of the input DNA (Folger et al., 1982; Thomas et al., 1986).

The infrequent isolation of head to head and tail to tail arrangements suggests that the predominant head to tail arrays are not built up by the random joining of DNA ends (Palmiter & Brinster, 1986). They seem to develop by some mechanism other than DNA ligation such as repeated cycles of reciprocal exchange. Microinjected or transfected linear DNA cannot form these concatemers by reciprocal exchange while it retains its original form. Presumably some of it becomes circularised by the joining together of the ends of the same molecule. Head to tail arrays could then be formed by reciprocal exchange between two circular molecules or between one circular and one linear molecule. The linearisation of supercoiled molecules in the cell has been proposed as an essential step in the process leading to the integration of transfected DNA copies. Indeed supercoiled DNA has been shown to be converted to linear and open circle forms (Weintraub et al., 1986), and Folger et al., (1982) show that linear molecules yield a higher transformation efficiency (transformants per ug) than supercoiled molecules.

The productive reaarangement of transfected DNA can lead to the activation of gene sequences from a cellular promoter on integration.

If circularized DNA molecules are broken randomly, they will form a population of circularly permuted linear molecules. Reciprocal exchange between such circularly permuted molecules can also generate tandem arrays. The combination of end to end ligation of fragments of linearised plasmids and the formation of arrays by reciprocal exchange allows for rearrangement of the foreign DNA sequences. The potential for expression of the foreign gene sequence on integration is dependent on the point at which the DNA is linearised. A rearrangement which retains the coding region of the transfected gene can be expressed if it integrates next to a cellular promoter in an active chromosomal environment. These are the requisites for a so called "promoter trap". Most plasmids will integrate as tandem arrays with the coding sequences in the correct spatial organisation. In this case expression of the transfected gene will be initiated from the intrinsic promoter. Other rearrangements will be non-productive and will be silent. Hiller *et al.*, (1988) have proved that the activation at low frequency of a promoterless *tk* gene occurs by this mechanism of productive rearrangement followed by integration adjacent to an active cellular promoter.

The BS2 promoter is inactive in BHK fibroblast cells.

BHK*tk*$^-$ fibroblast cells were transfected with plasmids containing identical SV40 and *tk* sequences and 2.2 kb of the 5' region of either *Mup* gene BS6 or BS2. The transformation efficiency of a plasmid containing the *Mup* BS2 promoter sequence was half that of an identical plasmid containing the *Mup* BS6 promoter (Table 2). At first inspection this relationship could be indicative of the relative strength of each promoter (Gorman *et al.*, 1983; Spandidos & Wilkie, 1983; Rutherford & Nienhaus, 1987). However, when RNA from several clones stably transformed with the plasmid pSV*BS2*tkSupF was analysed by Northern blotting, the *tk* specific transcripts were found to be very heterogeneous in size. None of the transcripts are of a size consistent with accurate initiation from the *Mup* BS2 CAP site and termination at the *tk* polyadenylation signal (Figure 12).

S1 nuclease and primer extension analysis was performed on polyA$^+$ RNA from each of these clones. The results from the S1 nuclease protection assay are shorter than expected. The S1 nuclease data is effective in mapping a discontinuity in the foreign DNA sequence. This could be due to a rearrangement for example insertion or deletion of the transfected DNA or alternatively a transition to BHK chromosomal DNA. The primer extension assay establishes the true start of transcription and is consistent with the Northern blot data. None of the selected clones contains the BS2 CAP site and only one (clone ES9) produced any RNA transcribed from *Mup* sequences,

and this amounted to only 25 bp (Figure 18). It is concluded that the Mup BS2 promoter is not active in BHKtk$^-$ fibroblast cells. This is supported by a lack of transient expression of the plasmid pSVBS2tkSupF (Table 1A).

Tk sequences linked to an inactive BS2 promoter can be activated in the same way as promoterless tk genes.

On the basis of the work described by Hiller et al., (1988) it is most likely that the tk sequences in the selected clones are activated by cellular promoter elements following the integration of a productively rearranged copy of pSVBS2tkSupF rather than the insertion or deletion of DNA into the plasmid sequence. The S1, primer extension and Northern blot data all agree with this hypothesis and the presence of rearranged copies of the plasmid are shown by Southern blotting (Figure 20). RNA sequencing was employed to show that BHK chromosomal DNA was responsible for the transition observed by the S1 experiments. This did not yield a result.

In cells transfected with the plasmid pSVBS6tkSupF, the fidelity of tk expression from the Mup BS6 CAP site is maintained giving a 1.47 kb transcript (Ghazal, 1986; Figure 12). On transfection, supercoiled plasmids are linearised prior to integration to generate a series of circularly permuted linear molecules which form a concatemer of tandem arrays by reciprocal recombination. The integration of arrays of plasmid pSVBS6tkSupF will lead to active expression of tk from the BS6 CAP site. The Mup BS2 promoter is

inactive in BHK*tk*$^-$ cells even in combination with the SV40 enhancer. The integration of tandem arrays of plasmid pSV*BS2*tkSupF will not result in the expression of *tk* from the *Mup* BS2 CAP site. Cells transfected with pSV*BS2*tkSupF can only actively express *tk* when a productively rearranged *tk* gene is integrated adjacent to an active cellular promoter. The distance of the CAP site of the cellular promoter from the *tk* coding region is reflected in the length of the *tk* specific transcripts. This is shown by the Northern blot and primer extension analysis (Figures 12, 16 & 18).

The transformation efficiency of pSVBS6tkSupF is higher than that of pSVBS2tkSupF.

As a result of the *Mup* BS6 promoter being active and the BS2 promoter being inactive, identical linearisation events will result in a greater proportion of pSV*BS6*tkSupF integrants being able to express HSV thymidine kinase. This accounts for the greater transformation efficiency of pSV*BS6*tkSupF compared to pSV*BS2*tkSupF. These events are summarised below:

Structure of plasmid integrant	Integration at active promoter	Expression of *tk*	
		BS6	BS2
Integral	NO	YES	*
	YES	YES	*
Non productive rearrangement	NO	*	*
	YES	*	*
Productive rearrangement	NO	*	*
	YES	YES	YES

This visualisation of events allows the expression of *tk* from a cellular promoter in cells transfected with pSVBS6tkSupF. None of the selected clones show a major aberrant size message which is indicative of this event (Figure 12 & Table 3). It is likely that this is a bias in sampling a small number of clones. Northern blot analyses of clones HS2 and HS6 do show some very faint transcripts of aberrant size (Figure 12). In Northern blots a 340 bp PvuII-HindIII fragment of pSVBS2tkSupF which contains the SV40 promoter (Figure 9) and the plasmid minus the *tk* sequences does not cross-hybridise with any of these larger transcripts. This suggests that the extra long transcripts may be transcribed from the chromosome as a result of rearrangements although there is no evidence that they express an active *tk* gene. If the clones contain more than one copy of transfected plasmid it is likely that

at least one of these will be intact and express the tk gene from the BS6 CAP site to give the major 1.47 kb transcript seen.

Several reasons could account for the inactivity of the BS2 promoter.

The observed lack of expression of tk from the BS2 CAP site in BHK fibroblasts could suggest that an enhancer in the BS2 sequence is defective or that the promotional strength of the BS2 promoter is insufficient to produce enough TK to broach the threshold level for survival under HAT selection. Alternatively, a mutation may have disrupted a transcriptional activator protein binding site which results in a consequent lack of transcription. Several sequence anomalies of the BS2 sequence (compared to BS6) are observed which could support either of these hypotheses (Chapter 5). A further possibility is that BHK fibroblasts lack a trans-acting factor which is absolutely required for the initiation of transcription at the BS2 CAP site.

SV40 enhancer sequences may increase the transformation efficiency of plasmid pSVBS2tkSupF relative to pOtk.

The mechanism of activation of tk in cells transfected with the plasmids pOtk and pSVBS2tkSupF are concluded to be the same. Clones which have been transformed with pOtk show heterogeneity in the sizes of tk specific RNA transcripts, consistent with the activation of tk by a cellular promoter (Hiller et al., 1988;

Figure 13). No transient expression of the plasmid is observed. This is consistent with the absence of a promoter in the construct (Tables 1A & B). However, the transformation efficiency of pSVBS2tkSupF is approximately 12 times that of pOtk (Table 2). This apparent discrepancy could be due to the influence of the SV40 enhancer and early promoter sequences present in pSVBS2tkSupF. Enhancer sequences may directly or indirectly stimulate recombination perhaps by making the recombinant plasmid a better substrate for recombination or by bringing the recombinant plasmid into a region of high recombinational activity. Alternatively, enhancer sequences may stimulate DNA-mediated transformation by increasing the probability that a gene is active after its integration into the host chromosome. A gene with an associated enhancer may be active in a usually silent chromosomal environment. In the latter model enhancer sequences may function by altering local chromatin structure so as to facilitate transcription of an associated gene.

There is evidence to show that when SV40 sequences are present in a transfected plasmid they remain close to the tk sequences when expression is directed from a cellular promoter. Pulm & Knippers, (1985) transfected mouse fibroblast tk^- cells with a plasmid containing the SV40 enhancer linked to a promoterless tk gene and isolated 3 low copy number integrants. Maps of the integrants were constructed by blot hybridisation. Two of the integrants had a mosaic structure. This consisted of an intervening piece of cellular DNA separating the tk gene SV40 element from other sections of the

plasmid which also includes small parts of the SV40 enhancer. The integration of the SV40 sequences close to a cellular promoter may enhance transcription from the promoter so that enough gene product is produced to ensure survival in HAT medium. Murnane (1986) transfected a permanent human cell line with a plasmid containing the SV40 early promoter and a promoterless neo gene. Southern blot analysis of five G418 resistant subclones indicated that there were heterogeneous rearrangements in the region of the neo gene which were unique to each subclone. In one subclone the rearrangement was a tandem duplication that resulted in the relocation of the SV40 promoter 5' to the neo gene. There is no evidence that the SV40 enhancer does remain in the vicinity of the tk gene in clones stably transfected with pSVBS2tkSupF. Another hypothesis is that the mouse Mup sequences could promote integration.

The effect of the SV40 enhancer sequence on the expression of the transfected DNA.

The HSVtk gene can be accurately expressed from a BS6 Mup promoter in BHKtk$^-$ cells only when it is linked to an enhancer. In clones isolated after transfection with pSVBS6tkSupF it is not known whether all of the copies of the foreign gene retain the SV40 sequences. A proportion of active integrants may have lost the SV40 enhancer but may be activated by cellular enhancer sequences. Work carried out by Hamada (1986a; 1986b) is pertinent in this case. They describe an "enhancer trap" which hinges on the observation that the SV40 promoter is enhancer dependent. The

expression of chromosomally integrated enhancerless and enhancer-containing genes were analysed. An enhancerless plasmid containing the SV40 early promoter linked to a chloramphenicol acetyltransferase (CAT) gene was co- transfected with a cloned *tk* gene into murine *tk*$^-$ L cells. TK$^+$ colonies appeared at a frequency of $5x10^{-6}$. The CAT activity in TK$^+$ transformants showed that the enhancerless CAT gene was as actively expressed as an SV40 enhancer containing transformant. Furthermore polyA$^+$ CAT specific RNA transcripts were of a size consistent with initiation from the SV40 promoter. These results were consistent with the activation of transcription of the CAT gene from the SV40 promoter by an endogenous cellular enhancer. Using this selection system two gene activator elements were isolated from cloned human DNAs (Hamada, 1986b). They show functional and structural properties characteristic of other enhancers (Chapter 5).

Ghazal (1986) found that in cells transfected with pSV*BS6*tkSupF the *Mup* BS6 promoter was used in preference to the SV40 promoter. From the work described above it could be concluded that the SV40 sequences were absent and the *Mup* promoter was activated by endogenous enhancers. This does not seem to be the case as TK$^+$ colonies were isolated at a high frequency of $3x10^{-4}$ transformants/cell (Table 3).

NB [Frequency calculated at 182 colonies per $3x(2x10^5)$ cells.]

The level of HSVTK activity in each of the transfected clones is constant.

HSV thymidine kinase activity was measured in a number of clones stably transformed with a variety of plasmids (Table 3). Surprisingly, the level of activity was approximately the same in each transformant.

Pulm & Knippers (1985) found that a promoterless *tk* gene was as actively expressed as an intact *tk* gene when transfected into mouse L *tk*$^-$ cells. Hamada (1986a) showed that in the random isolation of enhancers from the human genome using an enhancer trap, surrogate enhancers were adopted resulting in a level of activity which was the same as that measured in cells containing the SV40 enhancer. Lang *et al.*, (1983) replaced the endogenous *tk* promoter with exogenous transcription control sequences from the human ε-globin and mouse β major globin genes and the SV40 and Moloney murine sarcoma viruses. Transfection of these constructs led to stable transformants with similar levels of *tk* activity. It is possible that only strong promoters, sufficient to raise *tk* expression above the required survival level, were selected in these experiments.

The 5' termini of tk specific RNAs could affect their translational efficiency.

The primer extension results show that clone ES6 has approximately

four times the amount of tk specific polyA$^+$ RNA as the other clones (Figure 16) although it does not have elevated TK activity relative to the other clones (Table 3). This suggests that TK activity is regulated by some post-transcriptional mechanism. Primer extension analysis has shown that the RNAs from each of the clones have different 5' termini. Differences in the translational initiation signals of the adopted promoters may result in RNA transcripts with different translational efficiencies. The efficiency of translation is also affected by the primary and secondary structures of the mRNAs in the region of the ribosome binding site. These are likely to be different for each of the RNA transcripts. Al Shawi et al., (1988) analysed the expression of a Mup-BS6/tk hybrid gene in transgenic mice in which tk expression is initiated from the Mup CAP site and consequently all tk specific transcripts have the same 5' terminus. The levels of tk mRNA and enzyme activity were found to be in broad agreement in different tissues and different lines.

The extent of HSVtk coding sequences within clones stably transfected with pSVBS2tkSupF.

Southern blot analysis of genomic DNA from individual cell clones cut with pairs of enzymes shows (by the criterion of fragment length) that some integrants have spatially intact Mup and tk sequences but that many rearrangements are also observed (Figure 20). Integrants containing intact Mup BS2 sequences are inactive. Of the remaining copies present in the clone probably only one needs

to be actively expressed from a cellular promoter to ensure survival of the cell under selective pressure. The analysis shows that there is no relationship between plasmid copy number and HSV*tk* activity (Figure 20 & Table 5). It is concluded that the selected clones contain a productively rearranged plasmid which expresses *tk* amongst a host of silent intact and nonproductive copies. Additional rearranged copies integrating next to weak promoters may result in the large faint *tk* specific transcripts present in the clones (Figure 12).

The first 100 base pairs of the tk coding region are NOT essential for survival in HAT.

Clone ES6 has two major *tk* specific transcripts (Figure 15 & 18). These are presumed to be messages from two fortuitous integrations following productive rearrangement of the transfected plasmid. The productively rearranged sequences which give rise to *tk* expression in clones ES3 and ES9 contain most of the *tk* coding region (Figures 17 & 18). Exceptions are clones ES2 and ES6 which lack approximately 100 and 80 bp of 5' *tk* coding region downstream of the translation initiation codon (Figure 17). This result shows that the first 100 bp of the *tk* coding region are not essential for survival in HAT medium. The *tk* protein translated from the short RNA in clone ES2 may have reduced activity. However its expression in sufficient quantities from an endogenous promoter could ensure survival and selection. Sequence analysis reveals the presence of 9 potential translation initiation codons in the first 400 bp of the

tk coding region (data not shown). Any of these could be used in the absence of the natural fMET codon. These results delineate an essential length of *tk* sequences which when translated allow the conversion of a cell from the TK^- to TK^+ phenotype. Analysis of the size of the protein would have to be undertaken to see which initiation site was actually being used. In the case of the clone ES6, which produces two *tk* specific transcripts, both messages could code for a functional thymidine kinase.

CHAPTER 11: ANALYSIS OF THE EXPRESSION OF THE Mup BS2 PROMOTER USING TRANSGENIC MICE.

Introduction.

Group 1 and group 2 *Mup* genes have evolved together in a 45 kb unit (Clark *et al.*, 1984b; Bishop, 1985) and can be distinguished on the basis of nucleic hybridisation results. Group 1 genes are classified as hepatic type genes (see Chapter 1) and although group 2 genes are linked to them in a 45 kb palindrome, there is no evidence to suggest that they are endogenously expressed in the liver. Indeed the *Mup* gene BS2 is presumed to be a pseudogene in the context of BS6 by virtue of lesions it contains in its protein coding sequence. Consistent with this observation, no BS2 specific transcripts are found on hybridisation of RNA from several *Mup* expressing tissues with a BS2 specific oligonucleotide probe (Shahan *et al.*, 1987b; McIntosh & Bishop, 1989). It is illogical to assume that the BS2 promoter sequence is unable to initiate transcription of heterologous coding sequences from the putative CAP site. Sequence analysis reveals that the BS2 promoter has all the essential *cis* components that define a eukaryotic promoter, including recognition sites for various nuclear binding factors. In addition, several features which may alter promoter strength and/or tissue specific expression relative to BS6 have been observed (Clark *et al.*, 1985; Chapter 5). The results (presented in this thesis) of transient and stable transfection assays with SV40 or *Mup*/*tk* hybrid genes, show that the BS2 promoter is not activated in

BHK*tk*$^-$ cells. As BHK fibroblasts do not endogenously express *Mup*, high level expression of BS2 may well require multiple tissue and hormone specific factors which are absent in fibroblasts.

The use of transgenic mice in gene expression studies.

In order to study regulated gene expression in an environment which is more compatible to that in which the endogenous gene is normally expressed transgenic animals are used. Gene expression in transgenic mice resembles the endogenous counterpart much more closely than that observed after gene transfer into tissue culture cells. Transgenic animals are a powerful experimental system in which to define the regulatory sequences involved in the differential expression of genes. The regulation of gene expression is investigated in transgenic mice by the gene transfer of specific DNA sequences which are carried in the mouse germ line. Transgenic mice have been useful for defining the *cis*- acting regulatory elements of a number of tissue specific and developmentally regulated genes including β-globulin, elastase, insulin and α-fetoprotein (Palmiter & Brinster, 1986; Cuthbertson & Klintworth, 1988; Gordon, 1989; Westphal, 1989).

Tissue specific expression of Mup genes and sexual dimorphism.

Pertinent to the work carried out in this thesis are two papers which deal specifically with the regulation of *Mup* gene expression in transgenic mice (Al Shawi *et al.*, 1988; Held *et al.*, 1989).

Al Shawi *et al.*, (1988) introduced a *Mup*/HSV thymidine kinase transgene containing 2.2 kb of group 1 gene BS6 5' promoter region into mice. Held *et al.*, (1989) used a *Mup* (BL6)/SV40 T-antigen hybrid containing 2.5 kb of 5' flanking region. Consistent with the classification of group 1 genes as hepatic type genes, expression of the reporter function was observed in the liver of transgenic offspring in both cases. This pattern of expression of the transgene shows that the sequences necessary for liver specific expression of BS6 and BL6 are located within the first 2.5 kb of DNA upstream of the CAP site. That the liver specific *cis*-acting sequences are located more than 350 bp away from the BS6 CAP site has been shown in transgenic mice harbouring a transgene containing only 350 bp of BS6 5' flanking sequence. Analysis of mRNA from various tissues and assays for thymidine kinase activity showed that liver expression had been abolished (Al Shawi & Bishop, personal communication).

Expression of the 2.2 kb *Mup*/HSV*tk* transgene (Al Shawi *et al.*, 1988) was lower in female than male liver in several lines and the same sex difference was observed in the rate of run-on transcription of the foreign gene in liver nuclei. When females of one of the sexually dimorphic lines were treated with testosterone, the levels of HSV*tk* RNA and TK activity were increased, although not to male levels. In these aspects of liver expression and lack of expression in 7 other tissues, this hybrid gene exhibits many of the characteristics of an endogenous group 1 gene.

A BS6/tk hybrid gene is misexpressed in the preputial gland and testis of transgenic mice.

Misexpression of the Mup/HSVtk transgenes containing either 2.2 kb or 350 bp of 5' BS6 flanking region was observed in the testis and the preputial gland (Al Shawi et al., 1988 & personal communication). Held et al., (1989) also observed expression of the BL6/SV40-T transgene in the preputial gland. The rat homologues to Mup, α2u-globulins, show a similar pattern of gene expresssion to endogenous mouse Mup genes (Laperche et al., 1983, MacInnes et al., 1986). One difference is that α2u-globulin is expressed at a high level in male and female preputial gland (Held & Gallagher, 1985). Evidence suggests that an evolutionary difference in cis-acting regulatory elements appears to be responsible for the synthesis of α2u-globulins in the rat preputial gland, in contrast to the lack of detectable Mup in the mouse preputial gland (Held & Gallagher, 1985). A rat α2u-globulin transgene was consistently expressed at very high levels in mouse preputial glands (da Costa Soares et al., 1987). This shows that the lack of expression of endogenous mouse Mup genes in the preputial gland cannot be explained by the absence of trans-activating factors. Misexpression could be attributed to silencer sequences which normally inhibit the expression of the endogenous Mup gene in the preputial gland and the testis, but which are located outside the promoter regions included in the hybrid constructs. Alternatively, a transgene might contain newly created sequences which activate expression in non-specific tissues. This

hypothesis is unlikely since the expression of both the BS6/tk and the BL6/SV40-T transgenes are observed in the preputial gland of different transgenic mice (Al Shawi et al., 1988; Held et al., 1989).

Misexpression of the BS6/tk gene in the testis and preputial gland are probably due to different causes.

Testicular misexpression could be due to the same cause as preputial misexpression. If this were the case, α2u-globulin expression in the rat testis might be expected. However, no α2u-globulin specific mRNA is detected in this tissue (Al Shawi et al., personal communication). This observation suggests either that misexpression of the hybrid Mup gene in the preputial gland and the testis has different causes, or that there is a common cause independent of the causes of α2u-globulin expression in the rat and mouse preputial glands.

Two observations (Al Shawi et al., 1988) suggest that different causes do bring about the misexpression of the hybrid gene in the preputial gland and testis. First, expression in the testis occurs in a transgenic line with no preputial gland expression. Secondly, tissue specific mRNAs differ in size. The absence of preputial gland expression in one line, 64 (Al Shawi et al., 1988), seems to be due to a position effect such that the expression of the gene is under the influence of flanking DNA sequences at the site of integration. This has been shown directly by cloning out the

transgene (using SupF) in line 64 and reintroducing it into recipient mice. Thymidine kinase assay analyses of the tissues of the resulting transgenic offspring show expression of the transgene in the preputial gland (Al Shawi et al., 1989). This is direct evidence of the effect of chromosomal position on transgene expression.

Testicular misexpression.

The actual cause of testicular misexpression remains unclear. However, the evidence implicates both the Mup promoter and the tk coding region as being important in determining this expression pattern. When present together, as in the BS6/tk transgene (with or without SupF), the expression of tk in the testis is very high and these mice are sterile (Al Shawi et al., 1988). When each component is present individually in a transgene, testicular expression of the reporter gene is observed at low levels for example Mup BS6/neo transgene (Jones, 1989 personal communication); Mup BS6/ α2u-globulin (Al Shawi, 1989). These mice are characteristically fertile. Some lines of mice transgenic for HSVtk linked to the mouse mammary tumour virus long terminal repeat expresses HSVtk in the testis (Ross & Salter, 1985) and are also sterile. In addition a HSVtk gene under the control of its own promoter was expressed in the testis (Kiessling et al., 1986). One of these mouse lines was fertile showing that the expression of tk in the testis does not invariably lead to sterility. In contrast a Mup(BL6)/SV40 T-antigen hybrid gene is not expressed in

the testis (Held et al., 1989). This could be because different Mup genes were used or that the SV40 T-antigen gene can not act in conjunction with the Mup promoter to initate transcription in the testis. There is evidence to show that the tk gene can influence the transcription of attached heterologous promoters. Koltunow et al., (1987) show that the efficiency of an avian keratin promoter could be increased about 20 fold in Xenopus oocytes when attached to the HSVtk gene.

HSVTK activity and sterility.

Available evidence suggests that a high level of HSV thymidine kinase activity in one or more testis cell type is the likely cause of male sterility (Al Shawi et al., 1989) although one exception has been documented (Kiessling et al., 1986). The enzyme could exert an effect in a number of ways which disrupt the cellular metabolism of nucleic acid precursors. These effects could stem from the expression of viral thymidine kinase at times in the cell cycle when when the cellular enzyme is normally repressed (Hofbauer et al., 1987) or from the enzyme having a different substrate specificity to cellular tk (Jamieson et al., 1974). In this respect HSVTK can efficiently phosphorylate deoxycytidine which can inhibit the normal pathway of pyrimidine phosphate metabolism (Kelly, 1983).

Expression of a Mup-BS2/HSVtk hybrid gene in transgenic mice.

RESULTS.

A 4485 bp HindIII-KpnI fragment containing approximately 2.2 kb of BS2 5' flanking region and the HSV thymidine kinase coding region was purified from the plasmid pSVBS2tkSupF; coordinates 2635-7120 (Figure 9). To facilitate the rescue of the hybrid gene from the genomic DNA of transgenic mice, the construct contains the bacterial amber suppressor tRNA gene SupF (Goldfarb et al., 1982). The DNA was microinjected into fertilised F_2 eggs derived from F_1 females [C57BL/6 x CBA/Ca] crossed with [C57BL/6 x CBA/Ca] F_1 males and the eggs transfered to the oviducts of pseudopregnant foster mothers. Of the pups born, 6 were identified as being transgenic: 5 males and 1 female. The males BST A and BST B were mosaics. They were fertile but did not transmit the transgene so no G_1 offspring were produced or lines established. The G_0 female transmitted the transgene at a frequency of about 50% to establish a line BST C. The males BST D, BST E and BST F and G_1 males from line BST C were all sterile (Table 6).

The copy number of the transgene in line BST C is 1-2 (data not shown). Keeping to the nomenclature in Al Shawi et al., (1988) a line is a group of male and female mice descended from the same G_0 female and containing the same transgenic insertion of foreign DNA.

Expression of the hybrid gene was studied by probing Northern blots with probes specific for HSV*tk* and by assaying HSV thymidine kinase activity in tissue homogenate supernatants. The presence of 0.4mM TTP in the assay has been shown to inhibit endogenous thymidine kinase activity by about 93-97%. The same amount of TTP inhibits HSV thymidine kinase by about 50% so that relatively low levels can be measured quite accurately (Jamieson & Subak-Sharpe, 1974; Al Shawi *et al.*, 1988).

In the following analysis the HSVTK activity data for mice BST D, E and F was compiled by Lisa Smith (PhD student).

HSVTK activity was detected consistently in the liver and testis of transgenic mice.

Foreign HSV*tk* mRNA and thymidine kinase activity was found consistently in the testis of *all* male G_0 mice and in male offspring of line BST C (Table 8). Expression consistently occurred in the livers of male and female mice from line BST C and the livers of BST D, BST E, BST F. In addition very low levels HSVTK were detected in the preputial glands of mice BST D and in the brain of mouse BST E (Table 7). No activity was observed in any of the mice in the kidney, salivary gland, lachrymal gland or muscle (Table 7). The amount of HSVTK expression in the liver is low for each integration, ranging from 0.24-0.6 U/mg protein. In comparison the associated testicular expression is very high, typically in a range of 37-83 U/mg protein (Table 8). In addition the enzyme activity in

Table 6.

Transmission of foreign genes from Go to G1 transgenics.

Go MICE	SEX	G1 OFFSPRING	
		TOTAL	TRANSGENIC
BST A	M	38	0
BST B	M	20	0
BST C (+ line)	F	17	8
BST D	M	0	0
BST E	M	0	0
BST F	M	0	0

Table 8.

HSV thymidine kinase activities in male tissues of transgenic mouse lines. Units (formation of 1pmol TMP/min) per milligram of protein; means and standard errors. Parentheses indicate number of mice used for analysis. All values are calculated over background (non transgenic litter mates) except for BST C.

Go LINE	LIVER	TESTIS	PREPUTIAL GLAND
BST A	0.019 (1)	4.47 (1)	0.066 (1)
BST B	0.017 (1)	23.3 (1)	0.025 (1)
* BST C	0.6 ± 0.045 (2)	83.0 ± 23.2 (2)	0.14 ± 0.06 (2)
BST D	0.55 (1)	37 (1)	0.37 (1)
BST E	0.24 (1)	86 (1)	0 (1)
BST F	0.3 (1)	67 (1)	0.03 (1)

* G1 mice.

Table 7.

A table to show the HSV thymidine kinase activities in tissues of transgenic mice, from line BST C, and control mice (non-transgenic litter mates).

Numbers are the mean HSVtk activity in the presence of 0.4mM TTP. Parentheses indicate number of mice used for analysis.

TISSUE	TRANSGENIC		NEGATIVE (controls) *	
	MALE (2)	FEMALE (2)	MALE (2)	FEMALE (2)
LIVER	0.60 ± 0.045	0.40 ± 0.08	0.04 ± 0.04	0.024 ± 0.005
TESTIS	83.0 ± 21.2		0.95 ± 0.9	
PREPUTIAL GLAND	0.14 ± 0.06	ND	0.20 ± 0.17	ND
KIDNEY	0.15 ± 0.1	0.11 ± 0.014	0.13 ± 0.03	0.14 ± 0.02
SALIVARY GLAND	0.04 ± 0.024	0.03 ± 0.007	0.05 ± 0.04	0.05 ± 0.03
LACHRYMAL GLAND	0.19 ± 0.13	0.04 ± 0.05	0.02 ± 0.13	0.05 ± 0.04
MUSCLE	0.07 ± 0.09	0.03 ± 0.03	0.08 ± 0.06	0.02 ± 0.02
BRAIN	0.36 ± 0.25	0.25 ± 0.08	0.40 ± 0.25	0.27 ± 0.05

* pmol(TMP)/min/mg at 37^{o}C; means and standard errors.

the livers of male and female mice from line BST C were approximately the same and do not appear to exhibit sexual dimorphism (Table 7).

Line-specific differences.

Two features of the expression of the foreign gene are line specific:

1) the occasional expression of the HSV*tk* gene in the brain in mouse BST E and in the preputial gland of mouse BST D.

2) The relative levels of expression of the transgene in different tissues (Table 8).

Sterility of males that carry the transgene.

The two fertile mice BST A & BST B were mosaic and did not transmit the transgene. The remaining G_0 males and male offspring from line BST C were sterile. All these mice show a high level of HSVTK activity in the testis (Table 8). Al Shawi *et al.*, (1988) show that the probable cause of sterilty in mice carrying a BS6/*tk* hybrid gene is a high level of thymidine kinase expression in the testis. These results are consistent with this theory. Female mice from line BST C were fertile.

Northern blot analysis of HSVtk specific transcripts in mice from line BST C.

Expression of the hybrid gene in mice of both sexes in line BST C was studied by probing Northern blots of liver and testicular RNA with an 840 bp HSV*tk* specific DNA fragment from pTK1 (Figure 21: samples B & D). RNA from the preputial gland, lachrymal gland, salivary gland, heart, muscle and brain was not analysed on the basis of the HSVTK enzyme assays. Clark *et al.*, (1985) define the group 1 BS6 CAP site by primer extension and SI mapping of *Mup* mRNA. Sequence alignment of the BS6 and BS2 genes positions the potential BS2 CAP site at 30±1 bp downstream of the TATA box (Figure 7). The expected size of a transcript running from the BS2 CAP site to the *tk* polyadenylation site is approximately 1.3 kb (1.45 with polyA). An RNA of approximately this size (1.42 kb) was detected in male liver RNA and in the testis RNA (1.3 kb; Table 9). The amount of this RNA is lower in the liver than in the testis (Figure 21: samples B & D). This is in agreement with the HSVTK assay. Each tissue contains a second smaller RNA of 0.9 kb. This reacts with the *tk* probe to the same extent as the 1.4 kb band in the liver and to a lesser extent to the 1.3 kb band in the testis. In addition the liver shows an additional faint transcript of 1.23 kb.

The smaller 0.9 kb band was presumed to be consistent with transcription initiation from cryptic start sites within the *tk* coding region, a phenomenon previously observed in cell lines

carrying integrated HSV tk genes (Roberts & Axel, 1982). However, a single stranded riboprobe specific for the sense RNA from the 5' tk coding region (Figure 14B & C; EcoRI-EcoRV) fails to hybridise to the 0.9 kb band in testicular RNA from a mouse line transgenic for a BS6/tk hybrid gene (C. Jones, personal communication). This supports the hypothesis that the 0.9 kb transcript is an antisense initiate. Probing of liver transcripts from this line, or testis and liver transcripts from line BST C have not yet been performed, but it is likely that the results would be consistent between different lines and tissues.

Comparison of the BS6/tk and BS2/tk hybrid gene transcripts.

Figure 21 shows an autoradiogram of a Northern blot of the tk specific transcripts from the testis and liver of a mouse from line 58 containing a BS6/tk transgene (Al Shawi et al., 1988) compared with those from a BS2/tk transgene. Table 9 shows the sizes of these transcripts as measured from the autoradiogram and Table 10 shows the mean HSV thymidine kinase activity in these tissues from each line.

Testicular transcripts. Tk specific transcripts in the testes of both BS6 and BS2 transgenic mice are identical in length running at 1.3 and 0.9 kb (samples A & B). The 1.3 kb band is consistent with initiation at the Mup promoter and termination at the tk polyadenylation site in each case. Primer extension and S1 analyses are currently being carried out to confirm this. The observed 0.9 kb

message is very likely to be an antisense transcript on the basis of recent work previously described (see above).

The intensity of the hybridising bands are approximately the same for the two transgenes and this is consistent with the similarities in the levels of HSV thymidine kinase activity measured (Table 10).

Liver transcripts. The expression of the two transgenes in the liver results in two identically sized transcripts of 1.42 and 0.9 kb (samples C & D). The 1.42 kb transcript is consistent with initiation at the Mup promoter and termination at the tk polyadenylation site and is the same size as the transcript intiating from the Mup BS6 CAP site in BHK fibroblasts (Table 3). Primer extension and SI mapping would be needed to confirm this. The difference in size between the transcripts observed in the testis and the liver may well reflect heterogeneity in the length of the polyA tail. The expression of the two transgenes seem to differ in the presence of an additional 1.34 kb transcript observed with BS6-tk (Sample C; Figure 21) and a 1.23 kb transcript with BS2-tk (Sample D). However, the small 1.23 kb transcript may also be presented in the BS6/tk sample but is obscured by the strong 1.42 kb signal. These additional transcripts possibly emanate from the Mup CAP site in each case, but differences in the length of the poly A tail results in the observed sizes. Another possibility is that the extra bands are antisense transcripts. Probing of the RNA with a sense strand-specific HSVtk riboprobe will help to confirm this hypothesis. Alternatively, extra bands may be a consequence of

Figure 21.

An autoradiogram of a Northern blot to show the sizes of HSVtk RNA in tissues from mice containing different MUP/tk transgenes.

Total cellular RNA from the testis and liver from an adult male mouse was applied to a 0.8% formamide gel and after blotting the filter was hybridised with a 840 bp PstI DNA fragment from pTKI.

TESTICULAR RNA.

Sample A: shows 1) 30ug and 2) 15 ug of testicular RNA from mouse line 58 containing a Mup BS6/tk transgene. 1 day exposure.

Sample B: as A except the RNA is taken from mouse line BST C containing a Mup BS2/tk transgene.

LIVER RNA.

Sample C: two exposure 1) 7 days 2)+3) 2 days of RNA from line 58 (male).

Sample D: two exposures 1) 7 days 2)+3) 4 days of RNA from line BST C (male). 25 ug of RNA was loaded in each case.

M = HSVtk specific markers (Figure 14B).

-ve = non transgenic testis RNA (line 58). -ve RNA samples from each tissue from each line were analysed. All were negative with respect to HSVtk specific RNAs.

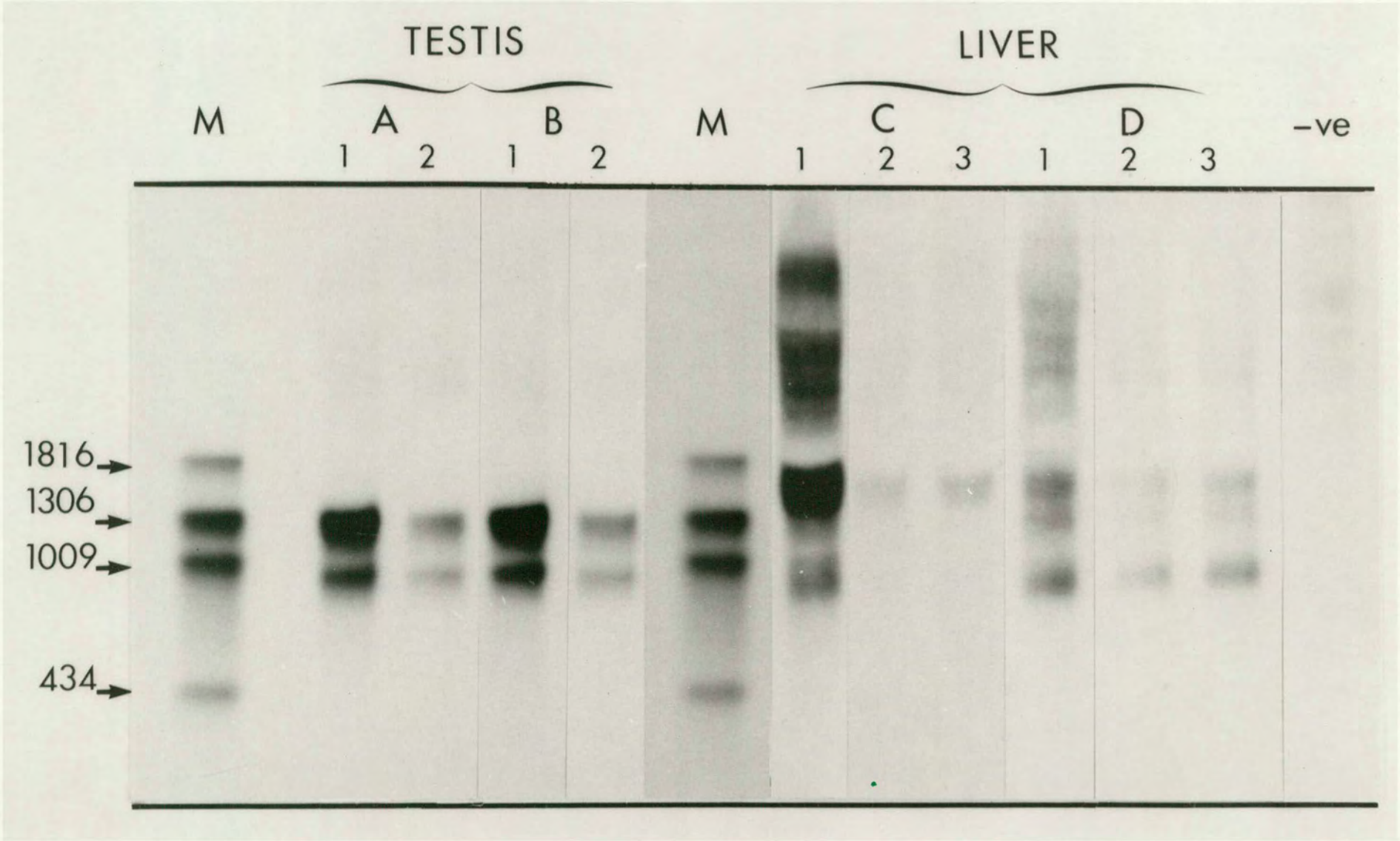

TESTIS
LIVER
M
A
B
M
C
D
-ve
1
2
1
2
1
2
3
1
2
3
1816
1306
1009
434

Table 9.

A table to show the sizes of HSVtk-specific transcripts in total cellular RNA extracted from tissues of transgenic mice.

TRANSGENE	EXPECTED SIZE (bp) to the *tk* poly-adenylation site.	ACTUAL SIZE OF TRANSCRIPT IN TRANSGENIC MICE (bp). TESTIS	LIVER
MUP-*BS6*-tkSupF	1276	1.3 0.92	1.42 1.34 0.92
MUP-*BS2*-tkSupF	1296	1.3 0.92	1.42 1.23 0.92

Table 10

To show the average levels of HSV thymidine kinase activity (pmol(TMP)/min/mg) means and standard errors in tissue samples from a mouse line containing either a BS6/*tk* (line 58) or a BS2/*tk* (line BST C) hybrid gene.

* Data taken from Al Shawi *et al.*, (1988).

	TRANSGENE	
	* MUP *BS6*-tkSupF	MUP *BS2*-tkSupF
TISSUE	LINE 58 (3)	LINE BST C (2)
TESTIS	31.0 ± 10	83.0 ± 21
LIVER	5.8 ± 1.15	0.6 ± 0.04
PREPUTIAL GLAND	9.9 ± 2.6	0

line specific position effects. Extra lines of transgenic mice will be required to resolve problem this by comparison.

The amount of tk-specific RNA in the liver of a mouse from line 58 is greater than that observed for line BST C (Figure 21). This is consistent with the approximate 10-fold difference detected in HSV thymidine kinase activity (Table 10).

Expression of the BS2/tk transgene in the preputial gland.

The expression of the BS2 gene in the preputial gland may be expected from the results of the work on the BS6/tk transgene. However, a level of HSVTK which was significantly different from background levels was only detected in the preputial gland of mouse BST D (Table 8). Thus it is difficult to conclude whether the transgene produces a very low basal level of expression in this tissue or none at all. With data for only one mouse, the expression observed for mouse BST D could be a position effect.

The patterns of expression in the liver, preputial gland and testis are line specific.

The level of expression and pattern of expression of a particular gene varies from one founder animal to another for example the expression of the BS2/tk transgene in the brain of mouse BST D. Line-specific effects are assumed to be due to the uniqueness of individual chromosomal integration events. Chromosomal position may

influence the accessibility of the genes to transcription factors. The quantitative analysis of transgene expression is complicated because multiple copies of the genes are usually integrated in a tandem array, and there is no means of determining how many of these genes are functional templates for transcription. The poor correlation between gene copy number and expression (Swift et al., 1984; Chada et al., 1985) suggests that only a few of the genes are expressed or that the entire array is sensitive to chromosomal position (Palmiter & Brinster, 1986).

DISCUSSION.

A 2.2 kb group 2 *Mup* gene promoter region and CAP site directs the transcription of the HSV*tk* reporter gene to the liver and the testis of transgenic mice. This observation is important because no endogenous BS2 specific gene transcripts have ever been observed (see Chapter 1; Shahan *et al.*, 1987b). The expression of the BS2/*tk* transgene in the liver enables one to classify the group 2 gene BS2 as a hepatic type gene.

The BS2 *promoter* contains no obvious structural deficiencies that would prevent the initiation of an mRNA at the probable CAP site, although several sequence anomalies exist which may contribute to transcriptional efficiency (Chapter 5). However, *tk* specific transcripts do not initiate from the BS2 CAP site when a BS2/*tk* construct is transfected into BHK*tk*$^-$ fibroblasts, even in the presence of the SV40 enhancer. Expression of *tk* results from the productive rearrangement and integration of the plasmid adjacent to an active cellular promoter (Chapter 10). In contrast, an identical plasmid carrying the BS6 *Mup* sequences expresses *tk* from the BS6 CAP site (Chapter 9). The fact that expression in the liver is observed in 4 different transgenic mice or lines rules out the possibility that *tk* expression in this tissue has resulted from a "promoter trap " situation.

There are several reasons why the BS2 promoter should initiate transcription in mouse cells but not in BHK fibroblasts.

The tissue culture evidence alone could suggest that the promotional strength of the BS2 promoter is insufficient to produce enough TK to broach the required threshold level for survival of a cell under HAT selection. Alternatively, BHK cells may lack tissue specific factors which are absolutely essential for expression from the BS2 CAP site for example enhancer binding factors. These factors would not be essential in the case of BS6 because of the intrinsic strength of its promoter. Other possibilities are: 1) that a mutation may have disrupted a regulatory protein binding site resulting in the lack of binding of a regulatory protein and a consequent lack of transcriptional initiation from the BS2 CAP site, or 2) that an enhancer in the BS2 sequence is defective.

The appropriate expression of the BS2/*tk* hybrid gene in transgenic mice rules out the last two possible explanations.

Why are no endogenous BS2 transcripts observed if the BS2 promoter is functional?

If, as seems apparent from the transgenic study, the BS2 promoter initiates accurate transcription in mouse cells, the question still arises as to why endogenous BS2 transcripts are never observed by oligo specific hybridisation (Shahan *et al.*, 1987b; McIntosh & Bishop, 1989). This may be because of the intrinsic instability of

the BS2 mRNA. Analysis of the BS2 coding region shows that the RNA sequence contains several lesions which render it untranslatable. Other acquired sequence mutations may cause mRNA destabilisation (see Chapter 5; Clark *et al.*, 1985b). If the *Mup* BS2 coding region can be replaced by a sequence that codes for a stable RNA which can be analysed on Northern blots, and a protein translated from this RNA which can be evaluated in biochemical assays for example HSV thymidine kinase, then the fidelity and tissue specificity of the BS2 promoter can be truely evaluated in transgenic mice. An alternative hypothesis is that the BS2 promoter is potentially functional, complete with tissue-specific regulatory elements, but is normally silenced as a consequence of its chromosomal position. This restraint is absent from the transgene. Interestingly, this type of "silencing" phenomenon has been demonstrated for other *Mup* genes (Shi *et al.*, 1989). They describe a *Mup* subfamily (*Mup*-1.5) containing two genes, *Mup*-1.5a and *Mup*-1.5b, that are nearly identical, differing at only three positions over the 4 kb transcription unit and 1 kb of flanking DNA. The similarity between these two genes extends over about 35 kb. Despite this extensive sequence homology (99.9%) only the 5a gene is expressed in BALB/c mice, primarily in the submaxillary gland, as shown by using specific oligonucleotide probes. However, when a 9.4 kb DNA fragment containing the 5*b* gene was introduced into the mouse germ line, mice in two of four transgenic lines expressed the gene at a high level and with the tissue specific characteristics of the 5a gene. These results suggest that the inactivity of the endogenous *Mup*-1.5*b* gene is

due to long range inhibitory position effects.

The BS2 gene is associated with the actively expressed BS6 gene in a 45 kb palindrome (Clarke et al., 1984b; Bishop, 1985). It is possible that the BS6 sequences in the vicinity of the BS2 gene confer on it a chromatin configuration which is incompatible with transcription.

Silent genes.

Silent genes of this type may be a general phenomenon. In the human α-globin gene complex, the functional ε-gene is linked to a pseudogene which carries one identifiable amber mutation (Hill et al., 1985). In some individuals the amber mutation is corrected by interchromosomal gene conversion. The corrected gene is not expressed in vivo but can be transcribed when introduced into HeLa cells or Xenopus oocytes (da Costa Soares et al., 1987). Similarly, in the human growth hormone gene (hGH) family, there is a difference between the hGH-5-encoding gene and the cDNA sequence. The hGH-5 gene can be expressed in vitro but there is no evidence that it is expressed in vivo (Seeburg, 1982). However, neither of these genes has been shown to be expressed properly when moved to a new chromosomal environment, and thus the possibility remains that they are inactive due to a defect in a regulatory element, for example an enhancer, rather than because of a long range position effect.

CONCLUSION.

Regardless of the criteria which prevent the endogenous expression of the BS2 gene, these experiments show that the BS2 promoter is very weak in comparison to the BS6 promoter. This is shown by the reduced expression of *tk* from the BS2 promoter in the liver. The similarities in the levels of expression of *tk* in the testes of transgenic mice lines 58 and BST C suggest that the intrinsic strength of the BS2 promoter is not important in determining the pattern of expression in this tissue.

Abbreviations used in the text.

A	adenine
AMV	avian myeloblastosis virus
aprt	adenosylphosphoribosyltransferase
ATP	adenosine 5' triphosphate
bp	base pair
BHK	baby hamster kidney cells
BSA	Bovine serum albumin
Budr	5-bromodeoxyuridine
C	cytidine
CTF	cellular transcription factor
cDNA	DNA copy of RNA
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DMEN	Dulbecco's modified Eagles medium
DNA	deoxyribonucleic acid
DTT	dithiothreotol
EDTA	diaminoethanetetra-acetic acid
G	guanine
GRE	glucocorticoid responsive element
HAT	Hypoxanthine, aminopterin, thymidine.
HSV	Herpes simplex virus
kb	kilobase
LTR	long terminal repeat
MMTV	mouse mammary tumour virus
mRNA	messenger RNA

MUP	mouse urinary protein
Mup	mouse urinary protein gene
neo	neomycin gene
NF1	nuclear factor 1
PIPES	Piperazine-N,N'-bis[2-ethanesulfonic acid]
PPO	2,5-diphenyloxazole
POPOP	1,4-bis-2-(4-methyl-5-phenyloxazolyl)-benzine
poly $(A)^+$RNA	polyadenylated RNA
RNA	ribonucleic acid
S1	single strand specific nuclease
SDS	sodium dodecyl sulphate
Sp1	specific transcription factor 1
SV40	simian virus 40
T	thymidine
tk	thymidine kinase (gene)
TK	thymidine kinase (enzyme)
TCA	tris-[hydroxymethyl]-aminomethane
TMP	thymidine monophosphate
TTP	thymidine triphosphate
U	uridine
UAE	upstream activator element

REFERENCES

Al-Shawi, R., J.Burke, C.T. Jones, J.P. Simmons, and J.O. Bishop. 1988. A Mup promoter-thymidine kinase reporter gene shows relaxed tissue-specific expression and confers male sterility upon transgenic mice. Mol. Cell. Biol. 8, 11:4821-4828.

Al Shawi, R., P. Ghazal, and J.O. Bishop. 1989. J. Mol. Evol. in press.

Ali, S., and A.J. Clark. 1988. Characterisation of the gene encoding ovine β-lactoglobulin. J. Mol. Biol. 199:415-426.

Addison, W.R., and D.T.Kurtz. 1986. Nucleotide sequences required for the regulation of a rat α2u-globulin gene by glucocorticoids. Mol. Cell. Biol. 6:2334-2346.

Angel, P., I. Baumann, B. Stein, H. Delius, H.J. Rohmsdorf, and P. Herrlich. 1987. 12-O-tetradecanoyl-phorbol-13-acetate (TPA) induction of the human collagenase gene is mediated by an inducible enhancer element located in the 5' flanking region. Mol. Cell. Biol. 7:2256-2266.

Baker, C., and E.B. Ziff. 1981. Promoters and heterogenous 5' termini of the messenger RNAs of adenovirus seratype 2. J. Mol. Biol. 149:189-221.

Banerji, J., S. Rusconi, and W. Schaffner. 1981. Expression of a β-globin gene is enhanced by a remote SV40 DNA sequence. Cell 27: 299-308.

Barth, R.K., K.W. Gross, L.C.Gremke, and N.D.Hastie. 1982. Developmentally regulated mRNAs in mouse liver. Proc. Natl. Acad. Sci. USA. 79:500-504.

Beato, M. 1989. Gene regulation of steroid hormones. Cell 56:335-344.

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

Bernstein, J.A., J.M. Porter, and M.D. Challberg. 1986. Template requirements for in-vivo replication of adenovirus DNA. Mol. Cell Biol. 6:2115-2124.

Bienz, M., and H.R.B. Pelham. 1986. Heat shock regulatory elements function as a inducible enhancer in Xenopus hsp70 gene when linked to a heterologous promoter. Cell 45:753-760.

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

Bird, A.P. 1987. CpG islands as gene markers in the vertebrate

nucleus. Trends in Genetics 3:342-347.

Bishop, J.O., A.J. Clark, P.M. Clissold, S. Hainey, and U. Francke. 1982. Two main groups of mouse urinary protein genes, both largely located on chromosome 4. EMBO J. 1:615-620.

Bishop, J.O., G.G. Selman, J. Hickman, L. Black, R.D.P. Saunders, and A.J. Clark. 1985. The 45kb unit of major urinary protein gene organisation is a gigantic imperfect palindrome. Mol. Cell. Biol. 5:1591-1600.

Bradford, M. 1976. A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analyt. Biochem. 72:248-254.

Breathnach, R., and P. Chambon. 1981. Eucaryotic split genes. Ann. Rev. Biochem. 50:349-383.

Briggs, M.R., J.T. Kadonaga, S.P. Bell, and R. Tjian. 1986. Purification and biochemical characterisation of the promoter specific transcription factor Sp1. Science 234:47-52.

Brinster, R., Y. H. Chen, and M. Trumbauer. 1981. Somatic expression of Herpes thymidine kinase in mice following injection of a fusion gene into eggs. Cell 27:223-231.

Brinster, R.L., H.Y. Chen, M.E. Trumbauer, M.K. Yagle, and R.D. Palmiter. 1985. Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. Proc. Natl. Acad. Sci. USA. 82:4438-4442.

Cavallini, B., J. Huet, J.L. Plassat, A. Sentenac, J.M. Egly, and P. Chambon. 1988. A yeast activity can substitute for the HeLa cell TATA box factor. Nature 334:77-80.

Chada, K., J. Magram, K. Raphael, G. Radice, E. Lacy, and F. Constantini. 1985. Specific expression of a foreign β-globin gene in erythroid cells of transgenic mice. Nature 314:377-380.

Chang, C. 1988. Molecular cloning of the human and rat cDNA encoding androgen receptors. Science 240:324-326.

Charney, P., R. Treissman, P. Mellon, M. Chao, R. Axel, and T. Maniatis. 1984. Differences in human α and β globin genes expression in mouse erythroleukemia cells: the role of intragenic sequences. Cell 38:251-263.

Charney, P., P. Mellon, and T. Maniatis. 1985. Linker scanning mutagenesis of the 5' flanking region of the mouse β-major globin gene: sequence requirements for transcription in erythroid and non-erythroid cells. Mol. Cell. Biol. 5:1498-1511.

Chirgwin, J.M., A.E. Przybyla, R.J. Macdonald, and W.J. Rutter.

1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294-5299.

Chodosh, L.A., A.S. Baldwin, R.W. Carthew, and P.A. Sharp. 1988. Human CCAAT binding proteins have heterologous subunits. Cell 53:11- 24.

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

Clark, A.J., P.M. Clissold, and J.O. Bishop. 1982. Variation between mouse urinary protein genes isolated from a single inbred line. Gene 18:221-230.

Clark, A.J., P.M. Clissold, R.A. Shawi, P. Beattie, and J.O. Bishop. 1984a. Structure of mouse urinary protein genes: different splicing configurations in the 3' non coding region. EMBO J. 3:1045-1052.

Clark, A.J., J. Hickman, and J.O. Bishop. 1984b. A 45kb DNA domain with two divergently orientated genes is the unit of organisation of the murine major urinary protein genes. EMBO J. 3:2055-2064.

Clark, A.J., A. Chave-Cox, X. Ma, and J.O. Bishop. 1985a. Analysis of mouse major urinary protein genes: variation between the exonic sequences of group 1 genes and comparison with an active gene outwith group 1 both suggests that gene conversion has occurred between Mup genes. EMBO J. 4:3167-3171.

Clark, A.J., P. Ghazal, R.W. Bingham, D. Barret, and J.O. Bishop. 1985b. Sequence structures of a mouse major urinary protein gene and pseudogene compared. EMBO J. 4:3159-3165.

Clarke, J., and W. Herr. 1987. Activation of mutated SV40 enhancers by amplification of wild type enhancer elements. J. Virology. 61:3536-3541.

Clayton, D.E., and J.E. Darnell. 1983. Changes in liver-specific compared to common gene transcription during primary culture of mouse hepatocytes. Mol. Cell. Biol. 3:1552-1561.

Clissold, P.M., and J.O. Bishop. 1981. Molecular cloning of cDNA sequences transcribed from the mouse liver endoplasmic reticulum mRNA. Gene 15:225-235.

Clissold, P.M., and J.O. Bishop. 1982. Variation in mouse major urinary protein genes and the Mup gene products within and between inbred lines. Gene 18:211-220.

Clissold, P.M., S. Hainey, and J.O. Bishop. 1984. Messenger RNAs coding for mouse major urinary proteins are differentially induced by testosterone. Biochem. genet. 22:379-387.

Corssi, S. 1988. Induction of proto oncogene fos transcript

through the adenylate cyclase pathway: characterisation of c-amp-responsive element. Genes and Devel. 2:1529-1538.

Covarrubias, L., Y. Nishida, and B. Mintz. 1986. Early post-implantation embryo lethality due to DNA rearrangements in a transgenic mouse strain. Proc. Natl. Acad. Sci. USA. 83:6020-6024.

Covarrubias, L., Y. Nishida, M. Terao, P. D'Eustachio, and B. Mintz. 1987. Cellular DNA rearrangements and early developmental arrest caused by DNA insertion in transgenic mouse embryos. Mol. Cell. Biol. 7:2243-2247.

da Costa Soares, V., R.M. Gubits, P. Fiegelson, and F. Constantini. 1987. Tissue-specific and hormonally regulated expression of rat α-2u-globulin gene in transgenic mice. Mol. Cell. Biol. 7:3749-3758.

Cuthbertson, R. A., and G. K. Klintworth. 1988. Transgenic mice-a gold mine for furthering knowledge in pathobiology. Laboratory Investigation. 58 5:484-497.

Davidson, I., E. Fromental, P. Augereau, A. Wildeman, M. Zenke, and P. Chambon. 1986. Cell type specific protein binding to the enhancer of SV40 in nuclear extracts. Nature 323:544-548.

Denis, M., L. Poellinger, A.C. Wikstom, and J.A. Gustafsson. 1988. Requirement of hormone for thermal conversion of the glucocorticoid receptor to a DNA binding state. Nature 333:686-688.

Derman, E. 1981. Isolation of a cDNA clone for mouse urinary proteins. Age and sex related expression of mouse major urinary proteins is transcriptionally controlled. Proc. Natl. Acad. Sci. USA. 78:5425-5429.

De Villiers, J., and W. Schaffner. 1981. A small segment of polyoma virus DNA enhances the expression of a cloned β-globin gene over a distance of 1400 bp. Nucl. Acids Res. 9:6251-6264.

De Villiers, J., L. Olson, J. Banerji, and W. Schaffner. 1982. Studies of the transcriptional enhancer effect. Cold Spring Harbour Symp. Quant. Biol. 47.

Dierks, P., A. Van Ooyen, N. Mantei, and C. Weissman. 1981. DNA sequences preceding the rabbit β-globin gene are required for formation in mouse L cells of β-globin RNA with the correct 5' terminus. Proc. Natl. Acad. Sci. USA. 78:1411-1415.

Dierks, P., A. Van Ooyen, M.D. Cochran, C. Dobkin, J. Reiser, and C. Weissman. 1983. Three upstream regions from the CAP site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell 32:695-706.

Dorn, A., J. Ballekens, A. Stauls, C. Benoist, and D. Mathis. 1987. A multiplicity of CCAAT box-binding proteins. Cell 50:863-872.

Dynan, W.S., and R. Tjian. 1983a. Isolation of transcription factors that discriminate between different promoters recognised by RNA polymerase II. Cell 32:669-680.

Dynan, W.S., and R. Tjian. 1983b. The promoter specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell 35:79-87.

Dynan, W.S., S. Sazer, R. Tjian, and R.T. Schimke. 1986. The transcription factor Sp1 recognises a DNA sequence in mouse dihydrofolate reductase promoter. Nature 319:246-248.

Dynan, W.S. 1986. Promoters for housekeeping genes. Trends in Genetics. 2:196-199.

Dynan, W.S. 1989. Understanding the molecular mechanism by which methylation influences gene expression. Trends in Genetics. 5:35-36.

Elkareh, A., A.J.M. Murphy, T. Fitcher, A. Efstratiadis, and S. Silverstein. 1985. "Transactivation" control signals in the promoter of the Herpes virus thymidine kinase gene. Proc. Natl. Acad. Sci. USA. 82:1002-1006.

Evans, R.M. 1988. The steroid and thyroid hormone receptor superfamily. Science 240:889-895.

Evans, R.M., and S.M. Hollenberg. 1988. Multiple and cooperative trans-activating domains of the Human glucocorticoid receptor. Cell 55:899-906.

Everett, R.D., D. Batty, and P. Chambon. 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.

Feinberg, A.P., and B. Vogelstein. 1984. Oligo-labelling of DNA fragments. Analyt. Biochem. 137:266-267.

Finlayson, J.S., R. Asofsky, M.Potter, and C.C. Runner. 1965. Major urinary protein complex of normal mice: origin. Science 149:981-982.

Fire, A., M. Samuels, and P.A. Sharp. 1984. Interactions between RNA polymerase II factors and templates leading to accurate transcription. J. Biol. Chem. 259:2509-2516.

Folger, K.R., E.A. Wong, G. Wahl, and M.R. Capecchi. 1982. Patterns of integration of DNA microinjected into cultured mammalian cells: Evidence for homologous recombination between injected plasmid DNA molecules. Mol. Cell. Biol. 11:1372-1387.

Ghazal, P., A.J. Clark, and J.O. Bishop. 1985. Evolutionary amplification of a pseudogene. Proc. Natl. Acad. Sci. USA. 82:4182-4185.

Ghazal, P. 1986. PhD thesis: University of Edinburgh.

Ghosh, P.K., P.Lebowitz, R.J. Frisque, and Y.Gluzman. 1981. Identification of a promoter component involved in positioning the 5' terminus of the SV40 early mRNA's. Proc. Natl. Acad. Sci. USA. 78:100-104.

Gidoni,D., J.T. Kadonaga, P. Chambon, and R. Tjian. 1985. Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. Science 230:511-517.

Giniger, M., and M. Ptashne. 1987. Transcription in yeast activated by a putative amphipathic alpha helix linked to a DNA binding unit. Nature 330:670-672.

Goldfarb, M., K. Schimizu, M. Perucho, and M. Wigler. 1982. Isolation and preliminary characterisation of a human transforming gene from T24 bladder carcinoma cells. Nature 296:404-409.

Gordon, J. W. 1989. Transgenic animals. Int. Rev. of Cytology. 115:171-229.

Gorman, C., R. Padmanabhan, and B.H. Howard. 1983. High efficiency DNA-mediated transformation of primate cells. Science 221:551-553.

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

Graves, B.J., P.F. Johnson, and S.L. McKnight. 1986. Homologous recognition of a protein domain common to the MSV LTR and the HSVtk gene. Cell 44:565-576.

Green, S., and P. Chambon. 1987. Oestradiol induction of a glucocorticoid responsive gene by a chimeric receptor. Nature 325:75-78.

Green, S., and P. Chambon. 1988. Nuclear receptors enhance our understanding of transcription regulation. Trends in Genetics. 49, 11:309-313.

Grosschedl, R., and M. Birnsteil. 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in-vivo. Proc. Natl. Acad. Sci. USA. 77:1432-1436.

Grosschedl, R., and D. Baltimore. 1985. Cell type specificity of

immunoglobulin gene expression is regulated by at least three DNA sequence elements. Cell 41:885-897.

Grosveld, G.C., E. de Boer, C.K. Shewmaker, and R.A. Flavell. 1982. DNA sequences necessary for transcription of the rabbit β-globin gene in vivo. Nature 295:120-126.

Hamada, H. 1986a. Activation of an enhancerless gene by chromosomal integration. Mol. Cell. Biol. 6, 12:4179-4184.

Hamada, H. 1986b. Random isolation of gene activator elements from the human genome. Mol. Cell. Biol. 6, 12:4185-4194.

Hastie, N.D., and W.A.Held. 1978. Analysis of mRNA populations of cDNA-mRNA hybrid mediated inhibition of cell free protein synthesis. Proc. Natl. Acad. Sci. USA. 75:1217-1221.

Hastie, N.D., W.A. Held, and J.J. Toole. 1979. Multiple genes coding for the androgen regulated major urinary proteins of the mouse. Cell 17:449-457.

Held, W.A., and J.F. Gallagher. 1985. Rat α2u-globulin mRNA expression in the preputial gland. Biochem. Genet. 32:281-290.

Held, W.A., J.F. Gallagher, C.M. Hohman, N. Kuhn, B. Sampsell, and R. Hughes. 1987. Identification and characterisation of functional genes encoding the mouse urinary proteins. Mol. Cell. Biol. 7 :3705-3712.

Held, W.A., J.J. Mullins, N.J. Kuhn, J.F. Gallagher, G.D. Gu, and K.W. Gross. 1989. T antigen expression and tumorigenesis in transgenic mice containing a mouse urinary protein /SV40 T-antigen hybrid gene. EMBO. J. 8, 1:183-191.

Herr, W., and J. Clark. 1986. The SV40 enhancer is composed of multiple functional elements that can be compensated for by one another. Cell 45:461-470.

Hill, A. V. S., R. D. Nicolls, S.L. Theins, and D.R. Higgs. 1985. Recombination within the human embryonic ξ-globin locus: A common ξ-ξ chromosome produced by gene conversion of the ψ-ξ gene. Cell 42:809-819.

Hiller, S., M. Hengstler, M. Kunze, and R. Knippers. 1988. Insertional activation of a promoterless thymidine kinase gene. Mol. Cell. Biol. 8, 8:3298-3302.

Hiscott, J., D. Murphy, and V. Defendi. 1980. Amplification and rearrangement of integrated SV40 DNA sequences accompany the selection of anchorage independent transformed mouse cells. Cell 22:535-543.

Hofbauer, R., E. Muellner, C. Seiser, and E. Wintersberger. 1987.

Cell cycle regulated synthesis of stable mouse thymidine kinase mRNA is mediated by a sequence within the cDNA. Nucl. Acids. Res. 15:741-752.

Jamieson, A.T., J.H. Subak-Sharpe. 1974. Biochemical studies on the Herpes simplex virus specified deoxypyrimidine kinase activity. J. Gen. Virol. 24:465-480.

Jannson, J.O., T.R. Downs, W.G. Beamer, and L.A. Frohman. 1986. Receptor-associated resistance to growth hormone releasing factor in dwarf "little" mice. Science 232:511-512.

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

Jones, K.A., J. Kadonaga, P. Rosenfeld, P. Kelly, and R. Tjian. 1987. A cellular DNA binding protein that activates eukaryotic transcription and DNA replication. Cell 48:79-89.

Jones, N.C., P.W. Rigby, and E.B. Ziff. 1988. Trans-acting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumour viruses. Genes & Devel. 2:267-281.

Kadesch, T., and P. Berg. 1986. Effects of the position of the SV40 enhancer on the expression of multiple transcription units in a single plasmid. Mol. Cell. Biol. 6:2593-2601.

Kadonaga, J.T., K.R. Carner, S.R. Masiaz, and R. Tjian. 1987. Isolation of cDNA encoding transcriptional factor Sp1 and functional analysis of the DNA binding domain. Cell 51:1079-1090.

Kakidini, H., and M. Ptashne. 1988. Gal-4 activates gene expression in mammalian cells. Cell 52:161-167.

Kato, S., R.A. Anderson, and R.D. Camerini-Otero. 1986. Foreign DNA introduced by calcium phosphate is integrated into repetitive DNA elements of the mouse L cell genome. Mol. Cell. Biol. 6:1787-1795.

Kaye, J.S., S. Pratt-Kaye, M. Bellard, G. Dretzen, F. Bellard, and P. Chambon. 1986. Steroid hormone dependence of four DNAaseI hypersensitive regions within the 7 kb flanking region of the ovalbumin gene. EMBO J. 5:277-285.

Keegan, L., G. Gill, and M. Ptashne. 1986. Separation of DNA binding from the transcription activating function of a eukaryotic regulatory protein. Science 231:699-704.

Kelley, W. N. 1983. Hereditary oroctic aciduria, p.1202-1226. In J. B. Stanbury, J. B. Wyngarden, D. S. Fredrickson, J. L. Goldstein, and M. S. Brown (ed). The metabolic basis of inherited disease.

McGraw-Hill, New York.

Kieney, M.P., R. Lathe, and J.P. Lecoq. 1983. New versatile cloning and sequencing vectors based on bacteriophage M13. Gene 26:91-99.

Kiessling, U., K. Becher, M. Strauss, J. Schoeneich, and E. Giessler. 1986. Rescue of a tk plasmid from transgenic mice reveals its episomal transmission. Mol. Gen. Genet. 204:328-333.

Knopf, J.L., J.R. Gallagher, and W.A. Held. 1983. Differential multi hormonal regulation of mouse urinary protein gene family in the liver. Mol. Cell. Biol. 3:2232-2240.

Knowles, B.B., C.C. Howe, and D.P. Aden. 1980. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigens. Science 209:497-499.

Koltunow, A.M., K. Gregg, G.E. Rodgers. 1987. Promoter efficiency depends upon intragenic sequences. Nucl. Acids. Res. 15:7795-7809.

Kozak, M. 1984a. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucl. Acids. Res. 12:857-872.

Kuhn, N.J., M. Woodworth-Gutai, K.W. Gross, and W.A. Held. 1984. Subfamilies of the mouse major urinary protein multi gene family, sequence analysis of cDNA clones and differential regulation. Nucl. Acids. Res. 12:6073-6090.

Kurtz, D.T., and P. Fiegelson. 1978. Multihormonal control of the mRNA for the hepatic protein α 2u-globulin. Litwack G. (ed) Biochemical Actions of Hormones Vol 5:433-455.

Kurtz, D.T. 1981. Hormone inducibility of rat α 2u-globulin genes in transfected mouse cells. Nature 291:629-631.

Lang, J.A, N.M. Wilkie, and D.A. Spandidos. 1983. Characterisation of eukaryotic transcriptional control signals by assay of Herpes simplex virus type 1 thymidine kinase. J. Gen. Virol. 64:2679-2696.

Laperche, Y., K.R. Lynch, K.P. Dolan, and P. Fiegelson. 1983. Tissue specific control of α 2u expression: constitutive synthesis in the submaxillary gland. Cell 32:453-460.

Lech, K., K. Anderson, and R. Brent. 1988. DNA-bound FOS proteins activate transcription in yeast. Cell 52:179-184.

Lee, W., A. Haslinger, M. Karin, and R. Tjian. 1987a. Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. Nature

325:368-372.

Lee, W., P. Mitchell, and R. Tjian. 1987b. Purified transcription factor Ap-1 interacts with TPA inducible enhancer elements. Cell 49:741-752.

Lichtsteiner, S., J. Wuarin, and U. Schibler. 1987. The interplay of DNA binding proteins on the promoter of the mouse albumin gene. Cell 51:963-973.

Littlefield, J.W. 1964. Selection of hybrids from matings of fibroblasts in-vitro and their presumed recombinants. Science 145:709-710.

Ma, X. 1987. PhD thesis: University of Edinburgh.

MacInnes, J.I., E.S. Nozik, and D.T. Kurtz. 1986. Tissue-specific expression of the rat α2u-globulin gene family. Mol. Cell. Biol. 6:3563-3567.

Macpherson, I., and M. Stoker. 1962. Polyoma transformation of hamster cell clones: an investigation of genetic factors affecting cell competence. Virology 16:147-151.

Mahon, K.A., P.A. Overbeek, and H. Westphal. 1988. Prenatal lethality in a transgenic mouse line is the result of a chromosomal translocation. Proc. Natl. Acad. Sci. USA. 85:1165-1169.

Maniatis, T., S. Goodbourn, and J. Fischer. 1987. Regulation of inducible and tissue-specific gene expression. Science 236:1237-1244.

McIntosh, I., and J.O. Bishop. 1989. Differential expression in male and female mouse liver of very similar mRNAs specified by two group 1 major urinary protein genes. Mol. Cell. Biol. 9:2202-2207.

McKnight, S.L., and E.R. Gavis. 1980. Expression of the Herpes thymidine kinase gene in Xenopus laevis oocytes: an assay for the study of deletion mutants constructed in-vitro. Nucl. Acids Res. 8:5931-5948.

McKnight, S.L., E.R. Gavis, R. Kingsbury, and R. Axel. 1981. Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene. Identification of an upstream control region. Cell 25:385-398.

McKnight, S.L., and R. Kingsbury. 1983. Transcriptional control signals of a eukaryotic protein coding gene. Science 217:316-324.

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

Moncollin, V., N.G. Miyamotto, X.M. Zheng, and J.M. Egly. 1986. Purification of a factor specific for the upstream element of the adenovirus-2 major late promoter. EMBO J. 5:2577-2584.

Moreau, P., R. Hen, B. Wasylyk, R. Everett, M. Gouls, and P. Chambon. 1981. The SV40 72 bp repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucl. Acids. Res. 9:6047-6068.

Mulligan, R.C., and P. Berg. 1980. Expression of a bacterial gene in mammalian cells. Science 290:1422-1427.

Myers, R.M., K. Tilly, and T. Maniatis. 1986. Fine structure genetic analysis of a β-globin promoter. Science 232:613-618.

Nishizuka, Y. 1984. The role of protein kinase C in cell surface signal transduction and tumour promotion. Nature 308:693-698.

Nonaka, M., H. Kimura, Y.D. Yeul, S. Yokoyama, K. Nakayama, and M Takahashi. 1986. Identification of the 5' flanking regulatory region responsible for the difference in transcriptional control between mouse complement C4 and S1P genes. Proc. Natl. Acad. Sci. USA. 83:7883-7887.

Norstedt, G., and R.D. Palmiter. 1984. Secretory rhythm of growth hormone regulates sexual differentiation of mouse liver. Cell 36:805-812.

Novak, T.J., and E.V. Rothenberg. 1986. Differential transient and long term expression of DNA sequences introduced into T-lymphocyte lines. DNA. 5:439-451.

Ohanian, S.H., S.B. Taubman, and G.J. Thorbecke. 1969. Rates of albumin and transferrin synthesis in-vitro in rat hepatoma derived H411EC3 cells. J. Natl. Cancer Inst. 43:397-405.

Palmiter R. D., and R. L. Brinster. 1986. Germ line transformation of mice. Ann. Rev. Genetics. 20:465-499.

Pellicer, A., D. Robins, B. Wold, R. Sweet, J. Jackson, I. Lowry, J.M. Roberts, G.K. Sim, S. Silverstein, and R. Axel. 1980. Altering genotype and phenotype by DNA mediated gene transfer. Science 209:1419-1425.

Perucho, M., D. Hanahan, and M. Wigler. 1980. Genetic and physical linkage of exogenous sequences in transformed cells. Cell 22:309-312.

Pervaiz, S., and K. Brew. 1985. Homology of β-lactoglobulin, serum retinol binding protein and protein HC. Science 228:335-337.

Pinkeret, C.A., D.M. Ornitz, R.L. Brinster, and R.D. Palmiter. 1987. An albumin enhancer located 10 kb upstream functions along with its

promoter to direct efficient, liver-specific expression in transgenic mice. Genes & Devel. 1:268-276.

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

Ptashne, M. 1988. How eukaryotic transcriptional activators work. Nature 335:683-689.

Pulm, W., and R. Knippers. 1985. Transfection of mouse fibroblast cells with a promoterless HSVtk gene: number of integrated gene copies and structure of single and amplified gene sequences. Mol. Cell. Biol. 5:295-304.

Queen, C., and D. Baltimore. 1983. Immunoglobulin gene transcription is activated by downstream sequence elements. Cell 33:741-748.

Rauscher, F.J., L.C. Sambucetti, T. Curran, R. Distel, and B.M. Spiegelman. 1988. Common DNA binding site for FOS protein complexes and transcription factor AP-1. Cell 52:471-480.

Renkawitz, R., G. Schutz, D. Van der Ahe, and M. Beato. 1984. Sequences in the promoter region of the chicken lysozyme gene required for steroid regulation and receptor binding. Cell 34:503-510.

Robins, D., M. Ripley, A. Henderson, and R. Axel. 1981. Transforming DNA integrates into the host chromosome. Cell 23:29-39.

Roberts, J.M., and R. Axel. 1982. Gene amplification and gene correction in somatic cells. Cell 29:109-119.

Ross, S.R., and D. Solter. 1985. Glucocorticoid regulation of mouse mammary tumor virus sequences in transgenic mice. Proc. Natl. Acad. Sci. USA. 82:5880-5884.

Rutherford, T., and A.W. Nienhuis. 1987. Human globin gene promoter sequences are sufficient for specific expression of a hybrid gene transfected into tissue culture cells. Mol. Cell. Biol. 7:398-402.

Sabbah, M., G. Redeuilh, C. Secco, and E. E. Baulieu. 1987. The binding activity of estrogen receptor to DNA and heat shock protein is dependent on receptor bound metal. J. Biol. Chem. 262:8631-8635.

Sawyer, L. 1987. One fold among many. Nature 327:659-662.

Schaffner, W. 1989. How do different transcription factors binding the same DNA sequence sort out their jobs? Trends in Genetics. 5:37-39.

Schenk, T. 1981. Transcriptional control regions: nucleotide sequence requirements for initiation by RNA polymerase II and III. Current Topics in Microbial Immunology. 93:25-46.

Schirm, S., J. Jiricny, and W. Schaffner. 1987. The SV40 enhancer can be dissociated into multiple segments, each with a different cell type specificity. Genes & Devel. 1:65-74.

Seeburg, P.H. 1982. The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone. DNA.1:239-249.

Serfling , E., M. Jasin, and W. Schaffner. 1985. Enhancers and eukaryotic gene expression. Trends in Genetics. 1:224-230.

Shahan, K., and E. Derman. 1984. Tissue-specific expression of major urinary protein (Mup) genes in mice: characterisation of Mup mRNAs by restriction mapping of cDNA and by in-vitro translation. Mol. Cell. Biol. 4:2259-2265.

Shahan, K., M. Gilmartin, and E. Derman. 1987a. Nucleotide sequences of liver, lachrymal and submaxillary gland mouse major urinary protein mRNAs: mosaic structure and construction of panels of gene specific synthetic nucleotide probes. Mol. Cell. Biol. 7:1938-1946.

Shahan, K., M. Denaro, M. Gilmartin, Y. Shi, and E. Derman. 1987b. Expression of six mouse major urinary protein genes in the mammary, parotid, sub-lingual, sub-maxillary and lachrymal glands and in the liver. Mol. Cell. Biol. 7:1947-1954.

Sharp, J.A., M. Wagner, and W. Summers. 1983. Transcription of HSV genes in vitro overlap of a late promoter with the 3' end of the early thymidine kinase gene. J. Virology. 45:10-17.

Shaw, P.H., W.A. Held, and N.D. Hastie. 1983. The gene family for major urinary proteins: expression in several secretory tissues of the mouse. Cell 32:755-761.

Shi, Z.P., R. Lee, and R. Weinmann. 1986. Protein factor(s) binding independently to two different regions of the adenovirus 2 major late promoter. Nucl. Acids. Res. 14:3729-3744.

Shi, Y., H.J. Son, K. Shahan, M. Rodriguez, F. Constantini, and E. Derman. 1989. Silent genes in the mouse urinary protein gene family. Proc. Natl. Acad. Sci. USA. 86:4584-4588.

Sigler, P.B. 1988. Acid blobs and negative noodles. Nature 333:210-211.

Snyder, M., A.R. Buchman, and R.W. Davis. 1986. Bent DNA in yeast autonomously replicating sequence. Nature 324:87-89.

Sodroski, J.G., C.A. Rosen, and W.A. Haseltine. 1984. Trans-acting transcriptional activation of the LTR of Human T lymphotropic viruses in infected cells. Science 225:381-385.

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

Spandidos, D., and N.M. Wilkie. 1983. Host specificities of papillomavirus, moloney murine sarcoma virus and simian virus 40 enhancer sequences. EMBO J. 2:1193-1199.

Spandidos, D., and N.M. Wilkie. 1984. Transcription and translation: A practical approach. Edited by Hames, B.D., and S.J. Higgins. IRL Press, Oxford:1-48.

Spiegelberg, T., and J.O. Bishop. 1988. Tissue-specific expression in mouse hepatocytes cultured in growth-restricting medium. Mol. Cell. Biol. 8:3338-3344.

Stewart, M., D. Forrest, D. McFarlane, R. Onions, N. Wilkie, and J. Neil. 1986. Conservation of the c-myc coding sequence in transduced feline v-myc genes. Virology 154:121-134.

Stringer, J.R. 1982. DNA sequence homology and chromosomal deletion at a site of SV40 DNA integration. Nature 296:363-366.

Struhl, K. 1987. The DNA binding domains of the jun oncoprotein and yeast GCN4 transcriptional activator protein are functionally homologous. Cell 50:841-846.

Sturm, R., T. Baumruker, B.R. Franza, and W. Herr. 1987. An eleven kd HeLa octamer binding protein interacts differently with two separate octamer related sequences within the SV40 enhancer. Genes & Devel. 1:1147-1160.

Swift, G.H., R.E. Hammer, R.J. MacDonald, and R.L. Brinster. 1984. Tissue-specific expression of the rat pancreatic elastase 1 gene in transgenic mice. Cell 38:639-46.

Szoka, P., and K. Paigen. 1978. Regulation of mouse major urinary protein production by the Mup-a gene. Genetics. 90:597-612.

Thomas, K.R., K.R. Folger, and M.R. Capecchi. 1986. High frequent targeting of genes to specific sites in the mammalian genome. Cell 44:419-428.

Unterman, R.D., K.R. Lynch, H.J.L. Nakashi, K.P. Dalan, J.W. Hamilton, D.V. Cohen, and P. Fiegelson. 1981. Cloning and sequence of several α2u cDNAs. Proc. Natl. Acad. Sci. USA. 78:3478-3482.

Varmus, H.E. 1984. The molecular genetics of cellular oncogenes.

Ann. Rev. Genet. 18:553-612.

Wagner, M.J., J.A. Sharp, and W.C. Summers. 1981. Nucleotide sequence of the thymidine kinase gene of Herpes simplex virus type 1.J. Virol. 38:593-605.

Walker, M.D., T. Edlund, A.M. Baulet, and W.J. Rutter. 1983. Cell specific expression controlled by the 5' flanking region of insulin and chymotrypsin genes. Nature 306:557-561.

Wallenberg, J.C., A. Nepveu, and P. Chartrand. 1984. Random and non-random integration of a polyomavirus DNA molecule containing highly repetitive cellular sequences. J. Virol. 50:678-683.

Wallenberg, J.C., A. Nepveu, and P. Chartrand. 1987. Integration of a vector containing rodent repetitive elements in the rat genome. Nucl. Acids. Res. 15:7849-7863.

Wasylyk, B., C. Wasylyk, and P. Chambon. 1984. Short and long range activation by the SV40 enhancer. Nucl. Acid. Res. 12:5589-5608.

Webster, N.J.G., S. Green, J.R. Jin, and P. Chambon. 1988. The hormone binding domain of the oestrogen and glucocorticoid receptors contain an inducible transcription activation function. Cell 54:199-207.

Weintraub, H., P.F. Cheng, and C. Conrad. 1986. Expression of transfected DNA depends on DNA topology. Cell 46:115-122.

Westphal,H. 1989. Transgenic mammals and biotechnology. FASEB J.. 3:117-120.

Wilkie, N.M., J.W. Clements, W. Boll, N. Mantei, D. Lonsdale, and C. Weissman. 1979. Hybrid plasmids containing an active thymidine kinase gene of Herpes simplex virus type 1. Nucl. Acids. Res. 7:859-877.

Wilkie, T.W., and R.D. Palmiter. 1987. Analysis of the integrant in Myk-103 transgenic mice which fail to transmit the integrant. Mol. Cell. Biol. 7:1646-1655.

Wold, B., M. Wigler, E. Lacy, T. Maniatis, S. Silverstein, and R. Axel. 1979. Introduction and expression of a rabbit β-globin gene in mouse fibroblasts. Proc. Natl. Acad. Sci. USA. 76:5684-5690.

Woodworth, C.D., and H.C. Isom. 1987. Regulation of albumin gene expression in a series of rat hepatocyte cell lines immortalised by simian virus 40 and maintained in a chemically defined medium. Mol. Cell. Biol. 7:3740-3748.

Wu, H.M., and D.M. Crothers. 1984. The locus of sequence directed and protein induced DNA binding. Nature 308:509-513.

Yoshimura, F., B. Davison, and K. Chaffin. 1985. Murine leukemia virus LTR sequences can enhance gene activity in a cell type specific manner. Mol. Cell. Biol. 5:2832-2835.

Zenke, M., T. Grundstrom, H. Matthes, M. Wintzereith, C. Schatz, A. Wildeman, and P. Chambon. 1986. Multiple sequence motifs are involved in SV40 enhancer function. EMBO J. 5:378-397.

Zheng, X.M., V. Moncollin, J.M. Elgy, and P. Chambon. 1987. A general transcription factor forms a stable complex with RNA polymerase B. Cell 50:361-368.

Additional references.

Allison, L.A., M. Moyle, M. Shales, and C.J. Ingles. 1985. Extensive homology among the largest subunit of eukaryotic and prokaryotic RNA polymerase. Cell 42:599-610.

Anderson, J.E., M. Ptashne, and S.C. Harrison. 1987. Structure of the repressor-operator complex of bacteriophage 434. Nature 326:846- 852.

Berg, J.M. 1986. Potential metal binding domains in nucleic acid binding proteins. Science 232:485-487.

Berg, J.M. 1989. DNA binding specificity of steroid receptors. Cell 57:1065-1068.

Blochinger, K., R. Bodmer, J. Jack, and J.Y. Jan. 1988. Primary structure and expression of a product from cut, a locus involved in specifying sensory organ identity in Drosophila. Nature 333:629-635.

Buratowski, S., S. Holin, L. Guarente, and P.A. Sharp. 1989. Five intermediate complexes in transcription initiation by RNA polymerase II. Cell 56:549-561.

Corden, J.L., D.L. Cadena, J.M. Ahearn, and M.E. Dahmus. 1985. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. Proc. Natl. Acad. Sci. USA. 82:7934-7938.

Courey, A.J., and R. Tjian. 1988. Analysis of Sp1 in-vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell 55:887-898.

Desplan, C., J. Theis, and P.H. O'Farrell. 1988. The sequence specificity of homeodomain-DNA interaction. Cell 54:1081-1090.

Evans, R.M., and S.M. Hollenberg. 1988. Zinc fingers: Gilt by association. Cell 52:1-3.

Fairall, L., D. Rhodes, and A. Klug. 1986. Mapping of the sites of

protection on a 5S RNA gene by the Xenopus transcription factor IIIA. J. Mol. Biol. 192:577-591.

Godowski, P.J., D. Picard, and K.R. Yamamoto. 1988. Signal transduction and transcriptional regulation by glucocorticoid receptor-Lex A fusion proteins. Science 241:812-816.

Hanes, S.D., and R. Brent. 1989. DNA specificity of the Bicoid activator proteins is determined by homeodomain recognition helix residue 9. Cell 57:1275-1283.

Hoey, T., and M. Levine. 1988. Divergent homeobox proteins recognise similar DNA sequences in Drosophila. Nature 332:858- 861.

Hollenberg, S.M., and R.M. Evans. 1988. Multiple and cooperative trans-activation domains of the human glucocorticoid receptor. Cell 55:899-906.

Hope, I.A., and K. Struhl. 1986. Function dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell 46:885-894.

Klug, A., and D. Rhodes. 1987. "Zinc fingers": A novel protein motif for nucleic acid recognition. Trends in Biochemical Sci 12:464-469.

Laughon, A., and M.P. Scott. 1984. Sequence of a Drosophila segmentation gene: protein structure homology with DNA binding proteins. Nature 310:25-37.

Levine, M., and T. Hoey. 1988. Homeobox proteins as sequence specific transcription factors. Cell 55:537-540.

Ma, J., and M. Ptashne. 1987. Deletion analysis of Gal4 defines two transcriptional activating segments. Cell 48:847-853.

Otting, G., Y-Q. Qian, M. Muller, M. Affalter, W.J. Gehring, and K. Wuthrich. 1988. Secondary structure determination for the Antennapedia homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif. EMBO J. 7:4305-4309.

Pabo, C.O., and R.T. Sauer. 1984. Protein-DNA recognition. Ann. Rev. Biochem. 53:293-318.

Ptashne, M. 1988. How eukaryotic transcriptional activators work. Nature 335:683-689.

Sadowski, I, J. Ma, S. Triezenberg, and M. Ptashne. 1988. Gal4-VP16 is an unusually potent transcriptional activator. Nature 335:563-564.

Schneunly, S., A. Kuroiwa, P. Baumgartner, W.J. Gehring. 1986. Structural organisation and sequence of the homeotic gene

Antennapedia of Drosophila Melanogaster. EMBO J. 5:733-739.

Tora, L., J. White, C. Brou, D. Tasset, N. Webster, E. Scheer, and P. Chambon. 1989. The human estrogen receptor has two independent non-acidic transcriptional activation domains. Cell 59:477-487.

Treisman, J., P. Gonczy, M. Vashishtha, E. Harris, and C. Desplan. 1989. A single amino acid can determine the DNA binding specificity of homeodomain proteins. Cell 59:553-562.

Umesono, K., and R.M. Evans. 1989. Determinants of target gene specificity for steroid/thyroid hormone receptors. Cell 57:1139-1146.

Webster, N.J.G., S. Green, D. Tassett, M. Ponglikitmongkol, and P. Chambon. 1989. The transcriptional activation function located in the hormone-binding domain of the human oestrogen receptor is not encoded in a single exon. EMBO J. 8:1441-1446.

Wolberger, C., Y. Dong, M. Ptashne, and S.C. Harrison. 1988. Structure of a phage 434 Cro/DNA complex. Nature 335:789-795.

Zehring, W.A., J.M. Lee, J.R. Weeks, R.S. Jokers, and A.L. Greenleaf. 1988. The C-terminal repeat domain of RNA polymerase II largest subunit is essential in-vivo but is not required for accurate transcription initiation in-vitro. Proc. Natl. Acad. Sci. USA 85:3698-3702.