NOTICE

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

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

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

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

AVIS

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

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

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

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

Canada

TRANSCRIPTIONAL REGULATION OF THE MURINE PGK-1 GENE

By

Leslie Sutherland

A thesis submitted to
the School of Graduate Studies and Research
in partial fulfillment of the requirements for the degree
of
Doctor of Philosophy

Department of Biology
University of Ottawa

ISBN 0-315-93611-8

Canada

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

I declare that I am the sole author of this thesis.

I authorize the University of Ottawa to lend this thesis to individuals or other institutions for the purpose of scholarly research.

I further authorize the University of Ottawa to reproduce this thesis by photocopying or other means, in whole or in part, at the request of other institutions for the purpose of scholarly research.

Leslie Sutherland

Acknowledgements

My appreciation is extended to the following people for their contribution to this work: Dr. Michael McBurney, my supervisor, for making this all possible; Drs. Jim Neelin, Doug Johnson, John Bell and Balwant Tuana, members of my advisory committee, for their commitment to my project, and their helpful suggestions; Drs. Doug Gray and Barbara Vanderhyden, of the Cancer Research Group, for their advice throughout this work; Karen Jardine, Jane Craig, Ricardo Marius and Manon Dubé for technical assistance; Beth Mason, for administrative support and invaluable help in the production of this thesis; Dr. René St-Arnaud of the Shriners Hospital for Crippled Children in Montreal, and McGill University, for extracts, advice and encouragement; the Medical Research Council of Canada, for financial support.

In addition, I would like to extend a special thank-you to the following people, who made this challenging experience an enjoyable one: Giovanna Pari, for making my transition into the McBurney domain as smooth as possible; Christine Pratt, for helping me through some difficult technical times; Susan Fournier, for her support, and unfailing optimistic outlook; Beth Mason, for help with all things personal, administrative and computational, and without whom I could never have had the most *interesting* camping experience of my life; Karen Jardine, for introducing me into the world of constructs, patiently answering questions she already answered, and driving me home on cold winter days; Barbara Vanderhyden and Monica Bialik, for having such cheerful dispositions that even the dullest days were brightened; Josée Coulombe and Maria Jaramillo, for being both pleasant and helpful people to work with; and Ninan Abraham, Ricardo Marius, Manon Dubé, Elizabeth Douville, Helen Tai, Peter Duncan, Rubina Ismail, Marco Di Fruscio and Suzette Delaloge for their comradery.

Finally, a very special thank-you goes to my parents for giving me total freedom, and sacrificing theirs, for five long years. Your unwaivering support encouraged the fulfilment of a dream.

ABSTRACT

The gene encoding the glycolytic enzyme phosphoglycerate kinase is transcriptionally regulated at two levels. Expression of enzyme is related to the glycolytic activity of the cell, and is highest in transcriptionally active cells. Expression is also regulated by X chromosome inactivation, as the somatically expressed Pgk-1 gene is X-linked. The role of 5'-flanking cis-acting DNA elements and trans-acting factors in the regulation of Pgk-1 expresssion was examined.

The murine Pgk-1 gene contains an upstream activator sequence (UAS) in its 5'-flanking region. This region was found to be responsible for elevating transcription levels at least ten-fold above basal *Pgk-1* promoter levels in P19 embryonal carcinoma (EC) cells. Part of this activity was attributed to the R2 protein binding site, first identified by DNase 1 footprinting techniques. Mutation of the middle region of R2 resulted in a 5-fold reduction in expression of a *Pgk-1* driven construct in stable transfection experiments into P19 cells. It was also determined that the R2 site was not important for transcription in P19 cells induced to differentiate with retinoic acid (RA).

In undifferentiated P19 cells, another UAS protein binding site, R1, was identified by band shift analysis. R1 could not be detected by footprint analysis, suggesting that the affinity of binding at R1 was lower than at R2. The mutation of the R2 site did not abolish protein binding, which led to the hypothesis that multiple factors were binding the DNA at R2. R1 was also hypothesized to interact with multiple factors. However, fractionation of the P19 nuclear extract and use in band shift studies against the R1 DNA resulted in a single fraction with binding activity, suggesting a single R1 DNA binding

protein and a non-DNA binding component. This non-DNA binding component at R1 was found to be tissue- or species-specific. Southwestern analysis in conjunction with fractionation experiments suggested that one of the R2 DNA binding proteins was approximately 70 kD and that the R1 DNA binding protein was 120 kD.

Treatment of the P19 cells with RA led to a reduction in gene expression. Two days after exposure to the drug, the contribution to expression from the UAS was reduced by 50%, and four days after exposure the UAS no longer contributed to gene expression. Protein binding to the UAS was also altered after RA-treatment. A new site of protein interaction was detected in the distal region of the UAS, at R3, and binding at R1 was altered. There was, therefore, a correlation between protein interaction within the UAS and gene expression during differentiation.

The results presented in this thesis demonstrate that the regulation of PGK-1 occurs, at least in part, at the level of gene expression, and that the UAS has an important role to play in regulating expression levels of the gene during differentiation. The results also suggest that the transcriptional stimulatory activity of the UAS depends on higher-order interactions between multiple low affinity DNA binding proteins which change upon differentiation.

Table of Contents

Acknowledgements iv
Abstract v
Table of Contents vii
List of Figures x
List of Tables xii
Abbreviations xiii

CHAPTER 1 1

General Introduction 1
1.1 Introduction 1
1.2 Phosphoglycerate kinase 2
1.3 Transcription regulation of RNA polymerase II transcribed genes . . 5
1.3.1 Transcription control elements 5
1.3.2 Transcription factors 12
1.3.3 Mechanisms of action 28
1.4 X chromosome inactivation 34
1.5 Transcriptional regulation of PGK 36
1.6 P19 cells 46
1.7 Project objectives 47

CHAPTER 2 49

Materials and Methods 49
2.1 Culture Conditions for EC Cells 49
2.1.1 Routine Culture 49
2.1.2 Conditions for Differentiation 50
2.2 Transfection and Selection of EC Cells 50
2.3 Extracts 52
2.3.1 Total Nuclear Extract 53
2.3.2 Fractionated Nuclear Extract 55
2.4 DNA Templates 60
2.4.1 Oligonucleotides 60
2.4.2 Activator Fragments 62
2.4.3 Transfected DNA 62
2.4.4 Labelling 64
2.5 Mutagenesis 64

2.6 Plasmid DNA preparation 65
2.6.1 Small scale plasmid preparation 65
2.6.2 Large scale plasmid preparation 66
2.7 RNA Preparation and Analysis 67
2.8 Protein-DNA Binding Interactions 69
2.8.1 Band Shift Assay 69
2.8.2 Footprinting Assay 69

CHAPTER 3 74

The UAS in Undifferentiated Cells 74
3.1 Introduction 74
3.2 Results 76
3.2.1 Binding to the UAS in male and female cells 76
3.2.2 Multiple factors bind to the UAS 78
3.2.3 Sites of interaction 82
3.2.4 Binding activity at R1 and R2 differs 89
3.2.5 R2 mutation affects binding and expression 91
3.2.6 R1 binding is tissue- or species-specific 100
3.3 Conclusions and Discussion 109

CHAPTER 4 112

The UAS in Differentiated Cells 112
4.1 Introduction 112
4.2 Results 113
4.2.1 Expression from the UAS containing promoter is down-regulated in differentiated P19 cells 113
4.2.2 Nuclear extract from differentiated P19 cells contains novel DNA binding protein 116
4.2.3 R3 binding activity is distinct from proximal UAS binding activity 119
4.2.4 R3 and R1 are key sites in the developmental transcriptional regulation of *Pgk-1* 122
4.3 Conclusions and Discussion 129

CHAPTER 5 132
Discussion 132

APPENDIX I 137
Southwestern Analysis 137

APPENDIX II 144
Screening an Expression Library 144

APPENDIX III
Papers

REFERENCES

List of Figures

FIGURE 1: Upstream activator sequence of the Pgk-1 gene. 4

FIGURE 2: Model of TFIID complex interacting with DNA. 7

FIGURE 3: Model of general transcription factor binding to promoter DNA. . . 8

FIGURE 4: Model of the pre-initiation complex associated with the DNA. 25

FIGURE 5: Consensus binding sites and protein:DNA interaction on the human *Pgk-1* promoter. 37

FIGURE 6: Consensus binding sites and protein:DNA interactions on the murine *Pgk-1* promoter. 38

FIGURE 7: Protein:DNA interaction associated with the expressed and non-expressed state of the murine *Pgk-2* gene. 40

FIGURE 8: Functional elements within the yeast *PGK* upstream activator sequence. 41

FIGURE 9: Microscopic view of P19 cells during nuclear extract preparation. . 54

FIGURE 10: Separation of nuclear and nuclear-free cell extracts. 56

FIGURE 11: P19 nuclear extract visualized by SDS-PAGE. 57

FIGURE 12: RA-treated cells have a different population of DNA binding factors. 58

FIGURE 13: Comparison of UAS binding activity in P19, C86 and P10 cells 75

FIGURE 14: Stability of UAS transcription complex 77

FIGURE 15: Sites of protein:DNA interaction 79

FIGURE 16: Consensus protein binding sites within the proximal region of the UAS 80

FIGURE 17: R2 produces a footprint 82

FIGURE 18: R2 footprinted sequence 83

FIGURE 19: R1 and R2 binding activities 86

FIGURE 20: Binding activity at R1 and R2 differs 88

FIGURE 21: Sequence of the R2 mutation 89

FIGURE 22: R2 mutation leads to altered binding 90

FIGURE 23: DNase 1 footprint of mR2 region 92

FIGURE 24: mR2 footprinted sequence 93

FIGURE 25: Effect of R2 mutation on expression in P19 cells 95

FIGURE 26: Nf-1 transcription factors in P19 cells 98

FIGURE 27: Mutant R1 E-box affects binding in HeLa cells 100

FIGURE 28: P19 cell heparin-agarose column fractions 102

FIGURE 29: HeLa cell heparin-agarose column fractions 103

FIGURE 30: R1 binding activity is tissue- or species-specific 104

FIGURE 31: UAS activity in RA-treated P19 cells 111

FIGURE 32: Nuclear extracts from differentiated P19 cells contain novel DNA binding protein 113

FIGURE 33: R3 binding activity is distinct from proximal binding activity . . 117

FIGURE 34: R3 binding activity interferes with proximal binding activities in RA-treated P19 cells 119

FIGURE 35: R3 binding activity does not interfere with proximal binding activities 122

FIGURE 36: Involvement of R2 site in RA-treated P19 cells 124

FIGURE 37: Murine UAS protein binding sites in relation to the known human protein binding sites 126

FIGURE 38: Model 130

FIGURE 39: Southwestern analysis of P19 nuclear extract 137

FIGURE 40: Screening of lambda gt11 P19 cDNA expression library 148

List of Tables

1 Nucleotide sequence of oligonucleotides

2 Designation and description of DNA fragments routinely used in binding studies

Abbreviations

AC	activator core
β-gal	β-galactosidase
CAT	chloramphenicol acetyltransferase
CTF	CCAAT-box transcription factor
DMSO	dimethylsulfoxide
DTT	dithiothreitol
EC	embryonal carcinoma
EDTA	ethylene diamine tetraacetic acid
HD	homeodomain
HLH	helix-loop-helix
HMG	high mobility group
HSH	helix-span-helix
HTH	helix-turn-helix
Inr	initiator element
lacZ	gene encoding Escherichia coli β-galactosidase
LTR	long terminal repeat
MAR	matrix attachment region
MOPS	3-(N-morpholino)propane-sulfonic acid
ONPG	O-nitrophenyl-β-D-galactopyranoside
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline

PEG	polyethylene glycol
PMSF	phenyl methylsulfonyl fluoride
RA	retinoic acid
RAR	retinoic acid receptor
RXR	retinoid X receptor
SAR	scaffold attachment region
SDS	sodium dodecyl sulfate
SSC	sodium saline citrate
SSPE	sodium saline phosphate EDTA
TAF	TBP associated factor
TBE	Tris-borate EDTA
TBP	TATA-box binding protein
TE	10 mM Tris-HCl, pH 8.0, 1 mM EDTA
TR	thryroid hormone receptor
UAS	upstream activator sequence
USE	upstream sequence element
USF	upstream sequence factor
XCI	X chromosome inactivation
ZIP	leucine zipper

CHAPTER 1

General Introduction

1.1 Introduction

Phosphoglycerate kinase is a key enzyme in carbohydrate metabolism, being the first ATP generating enzyme in the glycolytic pathway. As such, the protein is essential to all cells, and expression of the gene is ubiquitous. The level of gene activity varies, however, depending on the metabolic activity of the tissue. In addition, the most widely expressed isozyme of phosphoglycerate kinase, PGK-1, is encoded by an X-linked gene which is subject to X chromosome inactivation in females. Therefore, pgk-1 gene expression is tissue-specifically and developmentally regulated.

The involvement of transcription factors in inducible, tissue-specific and developmentally regulated gene expression is well established, and in the mechanism of X chromosome inactivation is inferred (see below). The subject of this thesis concerns an examination of the role that transcription factors play in the developmental regulation of pgk-1 gene expression.

1.2 Phosphoglycerate kinase

Phosphoglycerate kinase is a 45 kD protein which catalyses the reversible conversion of 1,3-diphosphoglycerate to 1-phosphoglycerate (Kozak et al 1974; VandeBerg, 1985). There are three isozymes of phosphoglycerate kinase in mice: PGK-1^a, PGK-1^b, and PGK-2 (Pegoraro and Lee, 1978; VandeBerg, 1985). PGK-1 is encoded by a 21 kb gene with 11 exons and 10 introns. *Pgk-1^a* and *1^b* differ by a single base-pair in the coding region, which translates into a single amino acid difference (Boer et al 1990), and by seven base-pairs within the first 500 base-pairs of the promoter region. *Pgk-2* appears to have evolved from reverse transcribed *Pgk-1* cDNA, and contains a tissue-specific promoter (McCarrey and Thomas, 1987; Boer et al 1987; Gebara and McCarrey, 1992).

The gene encoding PGK-1 is located on the X chromosome and the gene encoding PGK-2 is autosomal. PGK-1 is expressed in all somatic cells, and in spermatocytes until the pachytene stage: PGK-2 is expressed only in spermatocytes, and is induced just prior to the cessation of PGK-1 transcription (McCarrey and Thomas, 1987; Tamaru et al 1990). *Pgk-1* is also subject to X chromosome inactivation. After about seven days of embryonic development the allele on one of the two X chromosomes in female cells becomes transcriptionally inert, only one *pgk-1* allele remaining active.

Although the gene encoding PGK-1 is constitutively expressed, except in sperm, the level of expression in any given tissue type varies depending on its glycolytic capability. Enzymatic activity is greatest in skeletal muscle, and decreases progressively from the liver, kidney, brain, heart, spleen and bladder to the lung (VandeBerg, 1985).

Transgenic mouse embryos which carry the bacterial gene encoding β-

galactosidase driven by the murine *Pgk-1* promoter, display extensive cell-to-cell variability of *Pgk-1* driven β-gal expression. Expression is highest in rapidly proliferating and metabolically active cells requiring high levels of glycolysis (McBurney et al 1993).

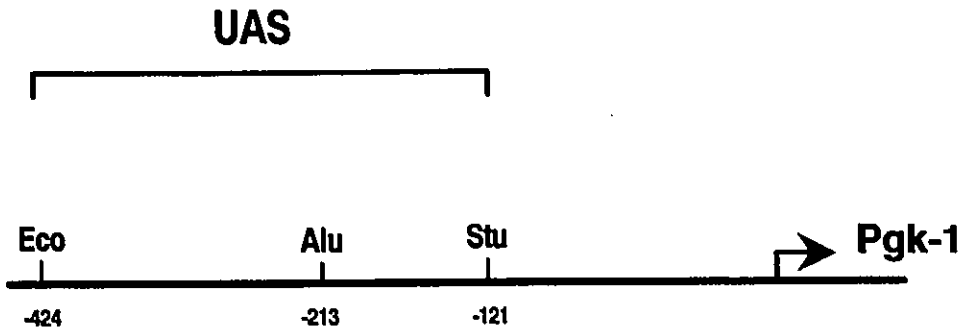
The requirement for high levels of phosphoglycerate kinase in actively cycling cells was observed by Kamel et al (1975) who noted that activity was higher in cancerous tissue than in normal tissue.

The level of PGK-1 expression has also been shown to vary during the course of development, in *Drosophila* (Roselli-Rehfuss et al 1992). Transcript levels were observed to rise during embryogenesis, remain high in the larva, fall in the pupa, and rise again in the adult, during emergence. This observation, along with those mentioned above, clearly demonstrates that PGK-1 expression changes during development and is greatest in cells with high metabolic activity.

The promoter of the gene encoding PGK-1 is very powerful when linked to transgenes and used in transfection experiments of cultured cells. Expression can be as high as that of the SV40 late and RSV promoters. Enhancement is due to a promoter specific upstream activator sequence (UAS) within the first 500 bp of the 5'-flanking region (McBurney et al 1991). The UAS sequence is broadly defined by the Eco R1 - Alu 1 and Alu 1 - Stu 1 restriction sites, a total of 300 bp (see Figure 1).

FIGURE 1: Upstream activator sequence of the Pgk-1 gene.

The upstream activator sequence (UAS) is encompassed by the Eco R1 and Stu 1 restriction sites, at -424 and -121 relative to the most 5'-transcription start site. DNA on both sides of the Alu 1 site is necessary for transcription elevation.


UAS
Eco
Alu
Stu
Pgk-1
-424
-213
-121

1.3 Transcription regulation of RNA polymerase II transcribed genes

1.3.1 Transcription control elements

The promoters of eukaryotic genes transcribed by RNA polymerase II are located immediately upstream from the start site of transcription and are typically about 100 base-pairs in length (Dynan and Tjian, 1985; McKnight and Tjian, 1986). This region of the gene not only efficiently directs initiation of transcription, but also stimulates transcription. Specific sequence motifs within the promoter define their mode of action. These are reviewed below.

The initiator element (Inr) is a pyrimidine-rich sequence [5'-PyAPyTCPyPyPy-3'] (Weis and Reinberg, 1992) which overlaps the start site of transcription (Nakatani, 1990), but is not present in all genes (Roy et al 1991). In the absence of other promoter elements, the Inr is the simplest functional promoter (Smale and Baltimore, 1989). The presence of either a TATA motif at position -25 relative to the transcription start site, or binding sites for transcription factor Sp1 (see below), augments Inr promoted transcription (Seto et al 1991). The Inr interacts with the transcription factors TFII-I (Roy et al 1991) or YY1 (Seto et al 1991; Shi et al 1991).

A short DNA sequence motif (the TATA-box), consisting of adenine and thymine bases, is located 25 to 30 base-pairs upstream of the transcription start site. The TATA-box is an important promoter component in determining the point of transcription

initiation (Müller et al 1988; Sawadogo and Sentenac, 1990; Wefald et al 1990; Conaway et al 1991). It has been proposed that functionally distinct TATA elements interact with different protein complexes (Simon et al 1988; Wefald et al 1990). The 38 kD TATA Binding Protein (TBP) binds to the TATA-box. In mammals, TBP binds to the TATA-box as a multi-subunit complex, termed TFIID.

TFIID contains TBP and at least seven other TBP Associated Factors (TAFs), ranging in size from 30-250 kD (see Figure 2) (Pugh and Tjian, 1992). The major TAFs which generally copurify with human TBP are $TAF_{II}250$, $TAF_{II}150$, $TAF_{II}110$, $TAF_{II}80$, $TAF_{II}60$, $TAF_{II}40$ and $TAF_{II}30$ (Chiang et al 1993). $hTAF_{II}250$ is the only TAF cloned, and is apparently identical to CCG1, a protein already cloned and implicated in cell-cycle progression (Hisatake et al 1993; Ruppert et al 1993; Weinzierl et al 1993). It contains multiple target sequences for various kinases, suggesting that it receives signals from various transcriptional regulators (Hisatake et al 1993). Timmers and Sharp (1991) have shown that TBP exists in HeLa cell extracts in at least two different multiprotein complexes, indicating that different subsets of TAFs associate with TBP.

A number of transcription factors have been identified which bind DNA around the TATA-box, including TFIIA, TFIIB, TFIIE, TFIIF, TFIIG/J and TFIIH (Gasch et al 1990; Wefald et al 1990) (see Figure 3). These factors are involved in promoting initiation, and bind in a cascade to form the pre-initiation complex (Zawel and Reinberg, 1992). This begins with TFIIA, which assists in the binding of TBP, but is not required by all promoters (Tyree et al 1993). TFIIB binds the DNA-bound TBP. In assays that monitor pre-initiation complex assembly, TFIIB binding can be a limiting step. Binding

FIGURE 2: Model of TFIID complex interacting with DNA.

TFIID is a protein complex consisting of the TATA-binding protein TBP and at least seven TBP-associated factors (TAFs). In mammals these are $TAF_{II}250$, $TAF_{II}150$, $TAF_{II}110$, $TAF_{II}80$, $TAF_{II}60$ $TAF_{II}40$ and $TAF_{II}30$. TBP binds to one side of the DNA and causes a local bend. The TAFs interact with the conserved domain of TBP, to one side of the DNA. Only $TAF_{II}250$ and $TAF_{II}150$ bind TBP directly. $TAF_{II}110$ binds $TAF_{II}250$. The placement of $TAF_{II}80$, $TAF_{II}60$, $TAF_{II}40$ and $TAF_{II}30$ within the complex is purely hypothetical. TBP protects bases from -31 to -25 (Buratowski et al 1988): the TFIID footprint extends from -45 to +35 (Sawadogo and Roeder, 1985; Zhou et al 1992). -45 to -10 are uniformly protected, while -10 to +35 is protected with a periodicity of 10 bp (Sawadogo and Roeder, 1985).

TFIID

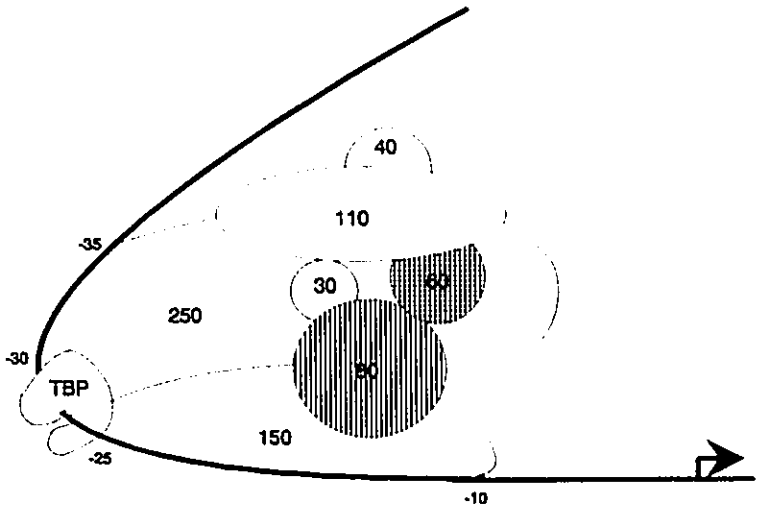
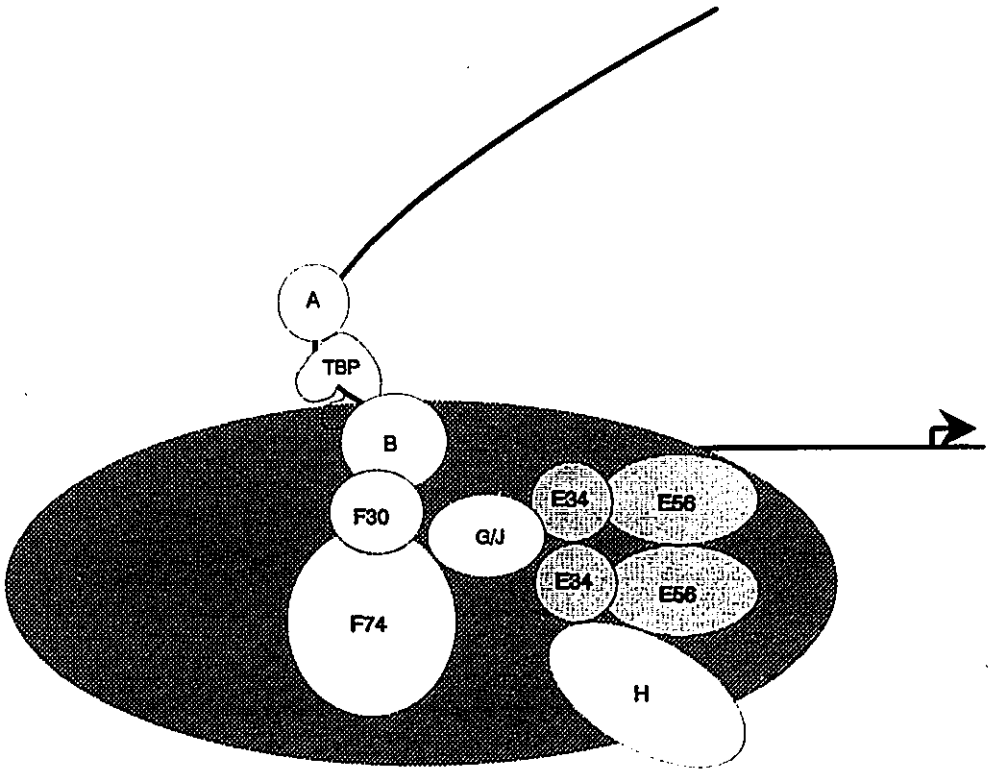


FIGURE 3: Model of general transcription factor binding to promoter DNA.

The general transcription factors which form part of the pre-initiation complex are TFIIA, TFIIB, TFIIE, TFIIF, TFIIG/J, TFIIH and TBP. TFIIA and TFIIB bind the DNA template, as well as TBP. The 30 kD subunit of TFIIF binds TFIIB. TFIIE and TFIIH bind one another. TFIIF is an RNA polymerase II associated factor. RNA polymerase II interacts with all the general transcription factors except TFIIA and TFIIG/J. Reviewed in Greenblatt (1992). The precise positioning of the general transcription factors is hypothetical.

General Transcription Factors



of TFIIB can be enhanced, however, through interactions with coactivator molecules (discussed later) (Lin and Green, 1991; Sundseth and Hansen, 1992). TFIIF, an RNA polymerase II associated protein, then escorts the non-phosphorylated form of RNA polymerase II to the transcription complex (Flores et al 1991). TFIIF is thought to have helicase activity (Lewin, 1990). Finally, the factors TFIIE (a kinase) (Peterson et al 1991), TFIIG/J (function unknown) and TFIIH (a multi-protein complex proposed to contain a helicase, a repair protein, XPBC/ERCC3 and a kinase involved in the phosphorylation of RNA polymerase II (Lu et al 1992; Peterson and Tjian, 1992; Schaeffer et al 1993) join and create a functional pre-initiation complex (Buratowski et al 1989; Inostroza et al 1991). Accurate basal level transcription can occur in the absence of the general transcription factors TFIIH, TFIIG/J, TFIIE and TFIIF(74), which apparently function to increase transcription levels (Tyree et al 1993).

TATA-less promoters also exist in a large variety of genes, and fall into two categories:(a) GC-rich (Sehgal et al 1988), and (b) non GC-rich. GC-rich promoters are characterized by multiple GC boxes which are the recognition sequence for the transcriptional activators Sp1 and ETF (Kageyama et al 1989). Genes which typically lack a TATA-box and are GC-rich are those which function constitutively, commonly referred to as housekeeping genes. In these genes transcription is generally initiated about 40 base-pairs downstream from the most proximal GC box, but can be spurious. On TATA-less, non-GC rich promoters, transcription depends largely on sequences near the start site of transcription and within the transcribed leader which contain binding sites for a large number of sequence specific DNA binding proteins (Smale and Baltimore, 1989;

Means and Farnham, 1990). On TATA-less promoters an Inr element may help to direct accurate initiation (Smale and Baltimore, 1989). On promoters that lack both a TATA-box and an Inr element transcription is frequently initiated at multiple sites.

The pre-initiation complex is also involved in transcription from TATA-less promoters (Pugh and Tjian, 1990; Dynlacht et al 1991; Zhou et al 1992). The Inr binding factor YY1, either alone (Roy et al 1991), or in conjunction with Sp1 (Zenzie-Gregory et al 1993), can anchor TBP to the DNA. The interaction of Sp1 with TBP appears to involve a tethering factor as Sp1 and TBP do not interact directly (Wiley et al 1992). This factor may be $TAF_{II}110$, which has recently been found to interact with Sp1 (Hoey et al 1993). Seto et al (1993) demonstrated that Sp1 and YY1 are also able to bind each other on the adeno-associated virus P5 promoter, but the significance of this finding remains to be determined. The indispensability of Sp1 to transcription is demonstrated by the fact that mutation of a single nucleotide in the GC-box in the promoter of the retinoblastoma gene prevents Sp1 from binding and leads to cancer in patients having this mutation (Sakai et al 1991).

The main difference between TATA-containing and TATA-less promoters may therefore be the mechanism by which they recruit TBP to the template and the affinity of their -30 binding regions for TFIID (Wiley et al 1992). Finally, although the TAFs are not necessary for basal level transcription on TATA-box promoters (Hoffman et al 1990; Peterson et al 1990), they are necessary for transcription on TATA-less promoters.

Sequence elements of 8 to 12 base-pairs, generally located between about -100 and -30 relative to the transcription start site, are termed upstream sequence elements

(USEs). They function in an orientation independent manner to stimulate basal transcription levels. Distance from the start site of transcription is critical for function. One common USE is the CCAAT-box, often found in a position 80-120 base-pairs upstream from the transcription start site, on either strand of the DNA. It is able to bind a variety of transcription factors including NF-1 (Jones et al 1987), NF-Y (Dorn et al 1987) and CBP (Hatamochi et al 1988).

Additional sequences, termed enhancers or silencers, are often associated with genes (Maniatis et al 1987; Guarente, 1988; Müller et al 1988; Sollner-Webb, 1988). They differ from USEs only in that they are distance independent, functioning to modulate transcriptional activity from distances of greater than 3000 base-pairs upstream of the transcription start site, from the promoter region, from within a gene, or from downstream of a gene (Müller et al 1988). Enhancers increase the rate of transcription by increasing the density of polymerase on the DNA template (Treisman and Maniatis, 1985; Weber and Schaffer, 1985; Li et al 1991). They function as tissue specific, inducible or temporal regulatory elements (Maniatis et al 1987). Genes may have multiple enhancers which function in response to more than one stimuli; for instance, immunoglobulin gene expression is tissue-specific (being restricted to lymphoid B cells), and temporally regulated (genes are only activated at a specific stage during lymphoid cell differentiation) (Maniatis et al 1987).

Repression of transcription may be as important as activation of transcription in the regulation of eukaryotic promoters. Repressive mechanisms are commonly divided into four categories (Levine and Manley, 1989). These are (i) direct repression by a

transcription factor, (ii) competition of an activator protein by a repressor protein for a common sequence, (iii) quenching of enhancer activity by binding of a repressor protein to a distinct but overlapping sequence, and (iv) squelching, by sequestration of activator proteins from the promoter.

Interestingly, the same factor can act as both a positive or negative regulator of transcription. This has been demonstrated for the papillomavirus E2 protein (Dostatni et al 1991), the lambda repressor (Ptashne et al 1980), the large T antigen of SV40 (Tjian, 1981), the engrailed homeo protein (Ohkuma et al 1990), and krüppel transcription factor (Sauer and Jäckle, 1993). The papillomavirus E2 protein acts as either a repressor or an activator depending on the position of its binding site relative to the basal promoter sequences. It represses transcription by interfering with the formation of the initiation complex. The further away from the initiation complex the E2 binding motif is, the more able E2 is to act as an activator of transcription. Krüpple, on the other hand, functions as an activator when present in low concentrations and bound to the DNA as a monomer. At higher concentrations a dimer of krüpple binds and acts as a repressor. Sauer and Jäckle (1993) propose that the monomer is able to function as an activator through interactions with components of the basal transcription complex, and that dimerization results in transcriptional repression by preventing or altering the basal factor interaction.

1.3.2 Transcription factors

As outlined above, discrete DNA sequence motifs are recognized by transcription factors whose functions are to regulate transcription. The motifs are many and varied, as are the

transcription factors with which they interact. The following is an overview of transcription factor structure and function.

1.3.2.1 Structure

The modular organization of transcription factors was first noted by Brent and Ptashne (1985). It is now well established that many transcription factors can be dissected into at least two functional domains: one is responsible for contact of the protein with the DNA and provides the specificity for the recognition of the target site; the second is involved in activation (Ptashne, 1988; Mitchell and Tjian, 1989; Struhl, 1989). Many transcription factors also possess dimerization domains, which modulate the activity of the protein via complex formation.

Protein-DNA binding

Although protein DNA-binding domains cover a variety of motifs, two of the most common are the zinc finger and a stretch of basic amino acids. These, and indeed all major known binding motifs, use α-helices for site specific recognition. In most cases, these binding domains make contact with the major groove of the DNA, and involve hydrogen bonds and van der Waals interactions. Interactions with the sugar-phosphate backbone can also be important for correct positioning of the protein in the DNA (Harrison, 1991).

Zinc finger binding domains are motifs within certain transcription factors (GCN4, Sp1, YY1) which interact with zinc to become a structural element capable of

binding specific DNA sequences. There are three classes of zinc containing DNA-binding proteins so far recognized, characterized according to the number and position of cysteine and histidine residues available for zinc coordination. Most proteins containing these modules have three or more fingers in direct succession. Sp1 contains three zinc finger domains (Kadonaga et al 1987). The fingers bind to the DNA at intervals of three base-pairs and lie in the major groove such that the protein wraps around the DNA. Although the zinc finger is necessary for binding, sequences surrounding the finger help to confer specificity (Pfeifer et al 1989).

A stretch of basic amino acids as a DNA-binding motif is common in DNA-binding proteins which dimerize, where the basic DNA-binding domain is frequently adjacent to a dimerization domain. Three common structural dimerization motifs present in DNA-binding proteins are the leucine zipper (ZIP), the helix-loop-helix (HLH), and the homeodomain (HD).

The basic leucine zipper motif (bZIP) gene family has been subdivided into several distinct classes on the basis of primary amino acid homology, DNA-binding specificity, and compatibility of the leucine zipper dimerization domain among members of each class. Included among these are the transcription factors Fos/Jun (Busch and Sassone-Corsi, 1990), CREB/ATF (Hai et al 1989; Ziff, 1990), C/EBP (Akira et al 1990) and TEF-DBP (Drolet et al 1991). When bound to DNA, the basic region becomes almost entirely α-helical. It recognizes sites 9-10 base-pairs long, with the consensus sequences having two-fold symmetry [AP-1: TGACTCA; CREB/ATF: TGACGTCA]. Contacts are made exclusively in the major groove and the protein wraps around the

DNA by about one half turn to either side of the centre of the site (Abel and Maniatis, 1989; Harrison, 1991). Some bZIP proteins, such as GCN4, are able to bind with nearly equal affinity to either the AP-1 or the CREB site (Harrison, 1991). It is interesting to note that, although the leucine zipper region is responsible for dimerization, if zipper residues are altered the proteins neither dimerize nor bind DNA (Abel and Maniatis, 1989).

The helix-loop-helix domain (Murre et al 1989b) and an adjacent basic region (Tapscott et al 1988) comprise the bHLH region of many transcription factors involved in development, the regulation of tissue-specific expression and sex-determination. The basic region, which is amino terminal to the HLH domain, is 10-20 amino acids in length and consists of 2-3 clusters of basic amino acid residues. It interacts with consensus sequences termed E-boxes [CANNTG], first defined within the immunoglobulin enhancers (Church et al 1985; Murre et al 1989a; Halazonetis and Kandil, 1991; Kerkhoff et al 1991). This binding activity may be mediated by other cellular factors (Thayer and Weintraub, 1993). The HLH domain consists of two segments capable of forming amphipathic α-helices (12-15 residues each) connected by a non-conserved "loop" region of variable length (9-20 residues) (Murre et al 1989b). The HLH element mediates dimerization. In almost all cases, a heterodimer of two such proteins is the active DNA binding species (Harrison, 1991).

Homeodomain proteins (HD) represent another family of DNA-binding proteins. The homeobox is a 180 base-pair protein-coding sequence found in nearly every eukaryote that has been investigated. These proteins are characterized by a helix-turn-

helix (HTH) DNA/protein binding domain. The DNA-binding sites of classic homeodomain proteins are generally A/T rich sequences (Rosenfeld, 1991), but many homeobox genes are able to bind degenerate sequences. A novel subclass of homeobox factors are the POU-domain containing proteins. Included in this group are the anterior pituitary specific factor Pit-1, the universal factor Oct-1, the B-cell specific factor Oct-2 and the regulator of cell fate in C. elegans Unc-86. In addition to the classic homeodomain, these factors have another conserved sequence termed the POU domain (Levine and Hoey, 1988). The region of the protein containing these two domains is approximately 150 amino acids in length and contains the POU-specific domain (POU-SD) and a POU-homeodomain (POU-HD) separated by a linker region of variable length (14-25 amino acids). The POU-HD corresponds to the classic homeodomain, with three predicted α-helical structures and a basic region at either end. The POU-SD contains two helices and a basic region at its N-terminal only. The POU-SD confers high affinity DNA binding and mediates protein-protein interaction, while the POU-HD is critical for DNA binding and critical for specific protein-protein interaction. It has been suggested that POU-domain proteins interact with their DNA recognition sites differently from homeodomain proteins, with both the POU-specific and POU-homeodomain contacting DNA. The best described sites for the POU-domain proteins Oct-1 and Pit-1 are variants of $(A/T)_{4-5}$TTTGCAT or $(A/T)_{4-5}$TATNCAT.

Ets domain proteins are a family of proteins characterized by a conserved DNA-binding region of approximately 85 amino acids (Nye et al 1992). Included in this family are Ets1, Ets2, Erg, Elk1, Elk2, Pu.1/Spi1 (Klemsz et al 1990), E74 (Burtis et al 1990),

Fli-1 (Ben-David et al 1991), gaba α (LaMarco et al 1991), Elf-1 (Thompson et al 1992), PEA3 (Xin et al 1992) and SAP-1 (Dalton and Treisman, 1992). Conservation of the ets domain sequence among family members ranges between 97% and 38% identity relative to Ets1, the founding member of the family. Within the 85 amino acid sequence is a basic region in the carboxy-terminal half, and a highly conserved, 10 amino acid leucine rich region in the amino-terminal half. All ets domain proteins contact the purine-rich sequence 5'-GGA(A)-3' (Karim et al 1990; Nye et al 1992; Gutman and Wasylyk, 1993). These sequences are found in the promoters or enhancers of various cellular and viral genes, including the murine sarcoma virus LTR, stromelysin, urokinase plasminogen receptor, interleukin 2 and Fos (Wasylyk et al 1990; Kola et al 1993). Ets1 and gaba α are known to bind DNA as monomers (Thompson et al 1991; Nye et al 1992). Each of the Ets family members has a different DNA-binding specificity, due most likely to recognition of unique core site flanking sequences and ets domain variability (Shyam et al 1990; Wasylyk et al 1992). Ets1 and Ets2 have a 76% homologous DNA-binding domain and are able to bind with high affinity to the same DNA motif (Wasylyk et al 1992). Even though the DNA-binding domain of the Erg gene is more homologous to chicken Ets-1 (c-Ets-1) and Ets-2 (~70%) (Rao et al 1987; Reddy et al 1987) than to E74 (~44%) (Burtis et al 1990), it shows more specific binding to E74 target sequences than to known target sequences of Ets1 and Ets2 (Shyam et al 1990). Elk-1 displays similar target specificity to Ets1 (Reddy and Rao, 1991). Binding specificity of the ets domain proteins may be further defined by interactions with other proteins, covalent modification, partitioning between cell compartments and cell specificity (He and

Rosenfeld, 1991).

The HMG domain is another 85 amino acid sequence, first identified in the "high mobility group" protein HMG1 (Walker et al 1980). It is present in many DNA-binding proteins, including transcription factors such as the T cell factor TCF-1 (van de Wetering et al 1991), the lymphoid enhancer-binding factor LEF-1 (Travis et al 1991), the testis-determining factor SRY (Sinclair et al 1990) and yeast mating-type proteins such as SteII (Kelly et al 1988; Staben and Yanofsky, 1991). The HMG domain appears to interact primarily with the minor groove of the DNA helix, with a few major groove contacts (Giese et al 1991), and induces bending of the DNA. The LEF-1 protein bends the DNA 130°, the SRY protein bends DNA 85°, and the SteII protein bends DNA 65° (Giese et al 1992).

Protein-protein binding

Transcription factors frequently, if not always, bind to DNA as dimers, heterodimers or multifactorial complexes. In some cases, protein is unable to bind its cognate DNA sequence unless it is pre-bound to another factor(s), as happens in the case of the bZIP family of proteins (Jones, 1990).

bZIP family members, including c-Jun, c-Fos and C/EBP, dimerize via a single amphipathic α-helix that has a heptad repeat of leucine residues on one face of the helix (Landschulz et al 1988). The dimerization of bZIP proteins is a very specific process. c-Fos, for instance, is unable to homodimerize, due to the structure of its leucine zipper, while c-Jun is able to bind DNA as both homo- or hetero-dimers: C/EBP forms only

homodimers (Neuberg et al 1989). Domain swap experiments of Kouzarides and Ziff (1989), and Sellers and Struhl (1989), have demonstrated that the resulting dimerization partners determine the specificity of DNA binding. As dimerization of bZIP proteins is necessary for DNA binding, loss of dimerization potential results in transcription inhibition. The binding activity of Fos and Jun is modulated by an inhibitor activity, IP-1 (Curran and Franza, 1988), which acts by competing for the leucine zipper domain (Auwerx and Sassone-Corsi, 1991).

The formation of dimers is also important to the bHLH family of transcriptional activator proteins, including the muscle specific factors, MyoD, Myf5 and myogenin, *Drosophila* genes involved in neurogenesis such as achaete-scute and daughterless, and genes expressed ubiquitously such as E2A (Williams and Tjian, 1991). Interestingly, some members of the bHLH family, including E2A and myogenin, also contain a leucine repeat, suggesting that interfamily dimerization may occur.

The bHLH family of transcription factors are divided into three classes (Murre et al 1989a; Barinaga, 1991). Class A comprises HLH factors which are ubiquitous. These include da (Caudy et al 1988; Cronmiller et al 1988), E2A (Murre et al 1989a), and hairy (Rushlow et al 1989). Class B HLH proteins are tissue restricted, and include AS-C (Villares and Cabrera, 1987) and the MyoD family of transcription factors, including Myf-5 and myogenin (Davis et al 1987; Braun et al 1989; Edmondson and Olson, 1989; Wright et al 1989). Class C HLH proteins are characterized by their inability to form heterodimers with either Class A or Class B proteins, which are able to form heterodimers with each other. Class C factors include Myc, USF (Gregor et al

1990), AP-4 (Hu et al 1990), TFEB (Carr and Sharp, 1990) and TFE3 (Beckmann et al 1990).

Class A factors can dimerize with Class B factors, such as MyoD and E12 (Davis et al 1990; Lassar et al 1991) or AS-c and da (Simpson, 1990). The Class C factor Myc is able to dimerize with Max, but not with factors in either of the other two classes (Blackwood and Eisenman, 1991). Myc and Max are structurally related in that both are not only HLH factors but also leucine zipper proteins, as are E12 (Murre et al 1989b), myogenin (Wright et al 1989), TFEB, AP-4, USF and TFE3 (Fisher et al 1992). Max is unable to interact with other bHLH, bZIP or bHLH.ZIP factors studied (Kouzarides and Ziff, 1988; Murre et al 1989a; Blackwood and Eisenman, 1991). Both the HLH and ZIP regions are required for protein-protein interaction; these two regions plus the basic region are required for DNA binding. TFE3 is unique in that it homodimerizes and is unable to form heterodimers even with other Class C members (Beckmann and Kadesch, 1991).

Interestingly, TBP has some similarity to Class C proteins. It has a highly conserved core domain at its 3'-end which has 80% homology between yeast and man. This conserved region constitutes roughly one half of the human protein and about 90% of the protein in *Arabidopsis*. This domain is the region which is necessary and sufficient for both DNA binding and transcription activity. It contains two direct repeats, both of which have similarity to helix II of the bHLH Class C family of proteins, specifically Myc. Whether or not this has implications for dimerization of bHLH factors with TBP is not known.

In cells that make all three classes of bHLH proteins, positive tissue specific regulators (MyoD, AS-C) compete with negative regulators (Id or emc) for binding to a ubiquitous protein (E12, E47 or da) (Benezra et al 1990; Sun et al 1991). The negative regulators Id1 and Id2 contain the HLH motif required for dimerization but not the basic amino acid region for DNA binding; consequently heterodimers of Id and any bHLH transcription factor are unable to bind DNA. Id1 and Id2 can inhibit Class A and Class B bHLH proteins but not those of Class C. emc (extramachrochaetae), the *Drosophila* homologue of Id, inhibits the function of the da and AS-C bHLH proteins involved in sex determination and neurogenesis (Ellis et al 1990; Garrell and Modolell, 1990).

Another DNA-binding and protein dimerization motif, the basic helix-span-helix (bHSH), has recently been identified in the transcription factor AP-2, a sequence-specific DNA-binding protein expressed in neural crest lineages and regulated by retinoic acid (Williams and Tjian, 1991). The AP-2 DNA-binding sequences do not resemble those from other transcription factors, but the overall organization is reminiscent of the bZIP and bHLH proteins where a basic region in association with an adjacent dimerization domain mediates stable DNA contact. The similarity in structure ends there, as the dimerization motif shares no homology with that of the HLH motif. The basic region covers 76 amino acids and the helix-span-helix covers 130 amino acids, 80 of which are in the span region. Homodimerization of this factor is essential for DNA binding. In fact, AP-2 exists as a dimer even in the absence of DNA.

The HTH motif of the homeodomain containing proteins and the Myb family of proteins (Frampton et al 1991) is responsible for dimerization. The POU proteins, which

contain this motif in addition to the POU-SD, exert either positive or negative transcriptional effects by binding to recognition elements as monomers or homodimers formed as a consequence of DNA-dependent cooperative interactions. Specific members of the POU family can also form heterodimers (Rosenfeld, 1991; Voss et al 1991). I-POU is a negative regulator of the POU-domain proteins. Like Id and emc (the negative regulators of the bHLH proteins), I-POU is able to dimerize with other family members but is unable to bind DNA (Foulkes et al 1991; Treacy et al 1991).

Physical associations also occur between distinct transcription factor families. Dimerization of the bZIP protein C/EBP and the Rel-family protein NF-κB leads to inhibition of promoters with κB binding elements, and the synergistic stimulation of promoters with C/EBP binding sites (Stein et al 1993). The bZIP protein Jun and the bHLH protein MyoD also dimerize, leading to the inhibition of myogenesis (Bengal et al 1992). Mutational analysis indicates that the zipper domain and the HLH domain are directly involved.

In addition to dimerization, some transcription factors are able to form large multiprotein complexes. This is the case for TBP in the TFIID complex, and the factors which interact with the serum responsive element of the c-Fos gene in proliferating rat liver. On the c-Fos serum response element, a complex of six proteins forms, three of which interact directly with the DNA and three of which appear to be involved via protein-protein interactions (De Belle et al 1991).

Activation

In addition to the DNA-binding and protein dimerization domains of transcription factors, is the activation domain. This part of the protein is specialized for protein:protein interaction with the transcription initiation complex, resulting in transcription activation (Hoey et al 1993). As is the case for DNA-binding domains (in the homeodomain proteins) and dimerization domains (in the bHLH.ZIP family of proteins), a transcription factor can have more than one activation domain. The transcription factor Sp1 has four activation domains (Courey and Tjian, 1988; Kadonaga et al 1988).

There are several chemically distinct types of activation domain (Mitchell and Tjian, 1989). Of the four Sp1 activation domains, two are glutamine rich, one is weakly basic, and the other exhibits no outstanding features or homologies. The CCAAT-box transcription factor (CTF) has a proline rich activation domain (Mermod et al 1989). The yeast transcription factors GCN4 and Gal4 have acidic activation domains (Gill and Ptashne, 1987).

TFE3, a murine transcription factor which binds to the enhancer of the immunoglobulin heavy chain gene, has a long form and a truncated form which lacks part of the activation domain and consequently acts as a trans-dominant negative modulator of transcriptional activity (Roman et al 1991).

Activation domains frequently interact with non-DNA-binding factors in their association with the initiation complex. These bridging molecules, termed adapters, coactivators, or mediators (Kelleher et al 1990; Pugh and Tjian, 1990; Flanagan et al 1991), act to bring disparate DNA-binding factors together (Lewin, 1990). $TAF_{II}110$ is

an example of a coactivator molecule. It serves as a bridge between the upstream activator protein Sp1 and TBP (Hoey et al 1993). As illustrated in Figure 4, $TAF_{II}110$ does not bind directly to TBP, but via $TAF_{II}250$ (Weinzierl et al 1993). The different TFIID complexes which exist (Timmers and Sharp, 1991) may result from the requirement for different co-activator molecules in association with TBP (Pugh and Tjian, 1991).

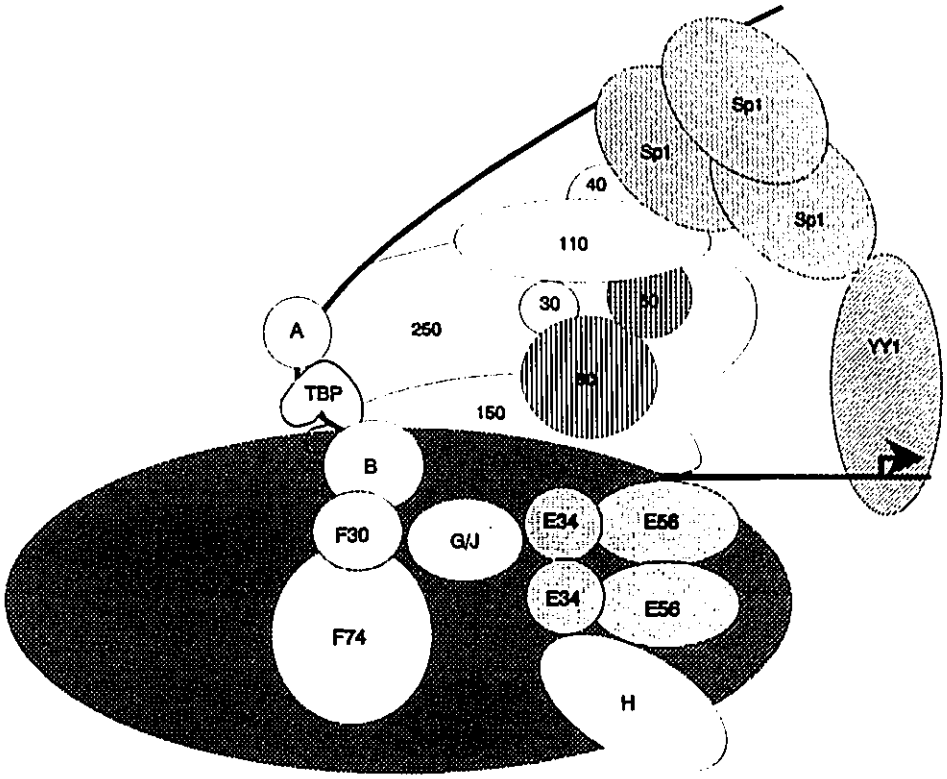
The Oct-1 transcription factor interacts with many different coactivator molecules, depending on the promoter context. For instance, on the 7SK promoter, Oct-1 interacts with the proximal sequence element transcription factor (Murphy et al 1992), on the IL2 promoter it interacts with an inducible factor which has not yet been identified (Ullman et al 1991), and on the herpes virus immediate early promoters Oct-1 binds with the viral protein VP16 (Gerster and Roeder, 1988). Other viral proteins which act as coactivators include E1a (Ptashne and Gann, 1990), Zta (Lieberman and Berk, 1991) and T-Ag (Gruda et al 1993).

Some interactions between upstream activator proteins and the pre-initiation complex occur via direct dimerization rather than with the aid of a bridging molecule. The Inr binding factor TFII-I interacts directly with the bHLH upstream activator protein USF (Roy et al 1991). The other known Inr binding factor, YY1, interacts directly with the upstream activator protein Sp1 (Lee et al 1993). TBP binds c-Rel, p65/RelA and dorsal (Kerr et al 1993), ATF (Horikoshi et al 1988) and the SRF (Zhu et al 1991). The general transcription factor TFIIB interacts with various upstream activator proteins including c-Rel (Kerr et al 1993), the fushi tarazu gene product (Colgan et al 1993) and c-Myc (Hateboer et al 1993).

FIGURE 4: Model of the pre-initiation complex associated with the DNA.

The pre-initiation complex is a large multifactorial complex including the general transcription factors and TFIID. In TATA-less promoters, the pre-initiation complex may include the upstream activator protein Sp1 and an Inr binding protein, e.g. YY1. For details of the binding mechanism refer to the text.

Pre-initiation Complex



When more than one Sp1 binding site is present, very high levels of Sp1 transcriptional activation can occur (50-80 fold greater than activation from a single site) (Carey et al 1990). One of the non-glutamine-rich activation domains is necessary for this synergistic activity, suggesting that it may be important for interaction between Sp1 molecules. Pascal and Tjian (1991) hypothesize that interaction of adjacent Sp1 complexes may generate a more effective activation surface to interface with components of the initiation complex. Increasing the number of Sp1 molecules in the complex, increases the number of activation domains.

Direct interaction between DNA-bound proteins is aided by the fact that many transcription factors bend the DNA when bound. Included in this group are TFIID (Horikoshi et al 1992), Jun and Fos (Kerppola and Curran, 1991), the estrogen receptor (Nardulli and Shapiro, 1992), the retinoid X receptor (Lu et al 1993), the thyroid hormone receptor (King et al 1993), NF-κB (Schreck et al 1990), heat shock transcription factor (Shuey and Parker, 1986), the POU proteins (Verrijzer et al 1991), the HMG proteins (Giese et al 1992; Lilley, 1992), the yeast multifunctional factor TUF (Vignais and Sentenac, 1989) and Myc/Max (Fisher et al 1992). The bend angles range from as little as 34° for the estrogen receptor to 130° for the LEF HMG protein. Each of the bHLH.ZIP family members bends the DNA with a similar angle and orientation (Fisher et al 1992). Multiple binding sites can increase or decrease the degree of bending (Ryder et al 1986; Shuey and Parker, 1986; Nardulli and Shapiro, 1992). Some DNA-binding proteins, including GCN4, NF-1 and the glucocorticoid receptor do not induce DNA bending (Gartenberg et al 1990; Luisi et al 1991; Verrijzer et al 1991).

The retinoid X receptor (RXR) will form homodimers, or heterodimers with either the retinoic acid receptors (RAR) or the thyroid hormone receptor (TR). Each of these dimers will bend the DNA toward the major groove but the degree of bending is dependent on the RXR dimerization partner and the structure of the DNA-binding site (Lu et al 1993). In this manner, variations in DNA bending, directed by different RXR binding partners, becomes a possible mechanism underlying the RXR protein transcriptional regulation of distinct retinoid and thyroid hormone responsive genes.

Kerppola and Curran (1991) studied DNA binding of the bZIP factors Fos and Jun. They found that interaction of Fos-Jun heterodimers, which have a much higher affinity for DNA than do Jun homodimers (Halazonetis and Kandil, 1991), interact with their consensus binding site in such a way as to bend the DNA toward the major groove, whereas Jun homodimers bend the DNA toward the minor groove. It was observed that the domains of both Fos and Jun responsible for dimerization and DNA binding will bend the DNA, but it is additional regions of these factors which influence the degree of bending. They speculated that aberrant expression of Fos and Jun regulated genes could result from the altered conformational changes to the DNA brought about by homodimer binding, leading to altered co-operative interactions with other trans-acting factors on cis-elements (Kerppola and Curran, 1991).

A more recent study by Kerppola and Curran (1993) examined the DNA-binding potential of a whole host of Fos and Jun related bZIP proteins which bind to the same DNA consensus site. They noted that DNA induced bending by combinations of the various family members ranged from 0 to greater than 40°C.

A role for DNA bending in transcription activation *in vivo* has only been inferred, not directly demonstrated (Nardulli and Shapiro, 1992). For instance, in the TCRα enhancer the LEF-1 binding site is 20 bp downstream from a ubiquitous CREB binding site and 20 bp upstream of a tissue-specific Ets binding site. A 130° bend in the DNA at the LEF-1 site could function by bringing the CREB and Ets proteins together to form a multiprotein complex. Altering the spacing between the CREB and LEF-1 sites does reduce transcription (Giese et al 1992).

A direct role for DNA bending in transcription repression has been suggested for the phage Ø29 p4 protein which bends DNA and represses transcription from the P_{A2b} promoter. Introduction of an intrinsically bent DNA sequence effectively mimics both the p4-induced DNA bending and the *in vitro* repression of P_{A2b} transcription (Rojo and Salas, 1991).

1.3.3 Mechanisms of action

The function of transcription factors is ultimately to alter chromatin structure for RNA polymerase entry. Current models of chromatin structure are reviewed in Getzenberg et al (1991).

Essentially, chromatin is based on a hierarchy of levels of DNA coiling. The first order of coiling is the 2 nm right-handed double helix, the second order is the nucleosome, and the third, the 30 nm filament.

The nucleosome consists of approximately 200 bp of DNA wound around a core octamer of histones H2A, H2B, H3 and H4. Histone H1 is associated with the

internucleosomal DNA. Six nucleosomes form one turn of a stacked spiral termed the 30 nm filament.

The 30 nm filament is anchored to the nuclear matrix (also termed nuclear scaffold) (see below) at discrete points termed MARs or SARs (for matrix attachment regions or scaffold attachment regions). This results in looping out of the intervening DNA. The loops average 60-90 Kbp in length and comprise anywhere from one to five genes (Goldman, 1988; van Driel et al 1991). Eighteen loops form one turn on the chromatid, termed a miniband. These minibands then stack along a central axis to form each chromatid (Getzenberg et al 1991).

MARs are AT-rich stretches of DNA from 200 to 1000 bp in length (Laemmli et al 1992; Boulikas, 1993). Originally identified in the *Drosophila* histone gene repeat (Mirkovitch et al 1984), MARs have since been found in the 5' and 3' flanking regions of many genes, often near enhancer elements (Cockerill and Garrard, 1986; Gasser and Laemmli, 1986; Stein et al 1991; Dworetzky et al 1992). Experiments involving cells in culture and transgenic animals have demonstrated that MAR sequences can act as transcriptional enhancers (Blasquez et al 1989; Stief et al 1989; Phi-Van et al 1990; Klehr et al 1991).

MARs bind to both the nuclear matrix and the nuclear lamina (Laemmli et al 1992). The nuclear matrix is a proteinaceous substructure composed of ubiquitous and tissue-specific proteins (Stuurman et al 1990; Romig et al 1992). Several components of the nuclear matrix (Berrios et al 1985; von Kries et al 1991), which are MAR-binding proteins, have been identified, including topoisomerase II (Adachi et al 1989), the

chicken attachment region binding protein ARBP (von Kries et al 1991), and the HeLa scaffold attachment factor SAF-A (Romig et al 1992). In addition, numerous transcription factors, including Myc, the large T antigen of SV40 virus, E1A from adenovirus, the steroid receptors in many estrogen and androgen responsive tissues (Getzenberg et al 1991; Stein et al 1991), and ATF (Dworetzky et al 1992), are associated with the nuclear matrix.

The nuclear lamina is another proteinaceous nuclear substructure with which MARs interact. It is a fibrillar network of intermediate filaments, which in mammals includes the lamins A, B_1, B_2 and C. The lamina lines the inner nuclear envelope and is believed to be a continuum of the intermediate filaments in the cytoskeleton (Gerace and Burke, 1988; Robson, 1989; Desser, 1990; Pienta et al 1991), and the nuclear matrix (van Driel et al 1991). It is this continuous matrix system which is hypothesized to enable the communication of signals from the exterior of the cell to the DNA, where appropriate modifications in organization are made which result in differences in gene expression (Getzenberg et al 1991). The lamins interact with both chromatin (Shoeman and Traub, 1990; Yuan et al 1991; Luderus et al 1992) and receptor molecules in the nuclear membrane (Foisner and Gerace, 1993). Recently, lamin B_1 was found to be a MAR binding protein (Luderus et al 1992). Investigations are ongoing regarding the possibility that the lamina-associated DNA is inactive (Luderus et al 1992).

DNA associated with the nuclear matrix is actively transcribed (Getzenberg et al 1991); however, genes in the looped out domain which constitutes the intervening DNA between MARs, are not transcribed (Stein et al 1991). Transcription which is stimulated

at a specific MAR is restricted to that loop domain, suggesting that the loop domain is a discrete transcription boundary (van Driel et al 1991; Laemmli et al 1992; Luderus et al 1992). Genes which are only expressed in specific cell types are only associated with the matrix in those cells. This differential organization of the DNA by the nuclear matrix, in a tissue specific manner, contributes to tissue-specific gene expression (Getzenberg et al 1991; Stein et al 1991). Unlike inert DNA, transcriptionally active DNA has a reduced, phosphorylated, H1 content (Huang and Cole, 1984; Karpov et al 1984; Schlissel and Brown, 1984; Hill et al 1990), histones H3 and H4 are highly acetylated (Vidali et al 1978; Hutcheon et al 1980), and H2A and H2B are ubiquitinated (Busch, 1984), a combination which results in a decrease in the linking number of the nucleosomes (Bradbury, 1992). In fact, nucleosomes are totally absent from the immediate vicinity of the site of transcription initiation (Reeves, 1984; Gross and Garrard, 1988; Tazi and Bird, 1990; Felsenfeld, 1992).

The chromatin switch model proposed by Laemmli et al (1992) states that histone H1 binds to MARs in vivo, and that other proteins have the ability to dislodge H1, resulting in a local chromatin opening (Felsenfeld, 1992). Additional events necessary for transcription activation then stabilize the active chromatin state. This model is supported by the following observations. First, certain activator proteins are able to recognize and bind to nucleosome containing DNA, and disrupt H1 mediated repression. The glucocorticoid receptor (Reik et al 1991; Rigaud et al 1991), Sp1 (Laybourn and Kadonaga, 1991), and GAL4 (Taylor et al 1991) are examples of such proteins. Transcription factor NF-1 further displaces the nucleosomes by functioning as a histone

acetylase (Oikarinen and Mannermaa, 1990). As acetylation of histone proteins leads to weaker DNA binding (Hebbes et al 1988; Megee et al 1990; Oliva et al 1990; Walker et al 1990; Braunstein et al 1993), this could be an important part of the transcription activation mechanism (Lee et al 1993). It is hypothesized that some of the initial binding transcription factors function as helicases, and disrupt the chromatin structure by generating a change in superhelical tension, which results in destabilization of the nucleosomes in the area. The change would then be propagated throughout the domain (Travers, 1992). In this manner other activator proteins which are unable to bind nucleosome containing DNA, such as the HSF (Taylor et al 1991) and TBP (Adams and Workman, 1993) would gain access to the promoter DNA. The DNA environment is then favourable for the general transcription factors and RNA polymerase II to enter and begin transcription. It is interesting to note that on some promoters the presence of nucleosomes actually increases transcription activity by facilitating the interaction of distally bound proteins (Schild et al 1993; Verdin et al 1993).

Transcription factors also play an important role in modulating transcription by acting as inhibitors or repressors. I have briefly mentioned some inhibitor proteins, such as I-POU, emc, and Id, which lack a DNA-binding domain but retain a dimerization domain. They function as repressors by heterodimerizing with activators, thereby reducing either their DNA-binding affinity or their activation potential. Some inhibitors of transcription possess DNA-binding domains but lack activation domains; they compete with activators for binding to the same site and thereby prevent activation. In addition, some repressors bind directly to activator proteins bound to the DNA and block their ability to transactivate.

p53, Dr1, Dr_2, NC1 and NC2 are repressors which repress transcription by binding directly to TBP (Meisteremst and Roeder, 1991; Liu et al 1993; Inostroza et al 1992; Seto et al 1992; Meisteremst et al 1991). Dr1, NC1 and NC2 interfere with the binding of TFIIA and TFIIB to the TBP protein. Human Dr_2, which has recently been identified as topoisomerase I, represses transcription in the absence of TFIIA: in the presence of TFIIA, repression by Dr_2 is overcome, and topoisomerase I activity leads to stimulation of transcription (Merino et al 1993). Dr_2 (and p53) only function as repressors on TATA-less promoters (Merino et al 1993; Mack et al 1993).

TEF is a repressor protein with homology to TBP. It is found in *Drosophila* embryos, in a highly restricted expression pattern. Like TBP, TEF binds the TATA-box, but unlike TBP, it neither binds TFIIB nor initiates transcription (Crowley et al 1993). It may therefore repress transcription by competing with TBP for binding to the TATA-box.

The retinoblastoma protein (Rb) also acts as a repressor of transcription. It controls cell proliferation by binding to certain transcription factors which would ordinarily bind TBP but are prevented from doing so by its intervention. These transcription factors include Myc, E2F and Pu.1 (Hateboer et al 1993; Hagemeier et al 1993).

The human thyroid hormone receptor β (hTHRβ) is another example of a transcriptional silencer. It acts by binding TFIIB, and presumably preventing further interactions of this factor with activator proteins. The presence of thyroid hormone, however, releases this interaction and results in activation of transcription (Baniahmad et al 1993).

1.4 X chromosome inactivation

X chromosome inactivation (XCI) is the process by which the majority of genes on one of the two X chromosomes in female placenta-bearing mammals becomes transcriptionally inert (Lyon, 1961; Gartler and Riggs, 1983; Grant and Chapman, 1988; Ballabio and Willard, 1992; Lyon, 1992). The process occurs between three and seven days post embryonic implantation (Monk and Harper, 1979) and is a heritable and reversible condition. In females harbouring more than two X chromosomes per diploid set, all but one will be inactivated (Endo et al 1982). The process is random in somatic cells (Endo et al 1982) but restricted to the paternally derived X chromosome in the extraembryonic endoderm and trophectoderm of the developing embryo (Takagi and Sasaki, 1975; West et al 1977). Inactivation originates at a single locus and spreads in *cis* in both directions along the chromosome, skipping some regions which are not inactivated (Rastan, 1983; Rastan and Robertson, 1985; Mohandas et al 1987). Transcription repression is followed by methylation, which is hypothesized to help in the maintenance of the inactive state (Lock et al 1987). In female germ cells there is reactivation of the inactive X chromosome prior to meiotic entry so that both X chromosomes are active during oogenesis (Gartler and Riggs, 1983; Grant and Chapman, 1988).

Despite the fact that its characteristics are well established, the mechanism of XCI is largely unknown. Of the many theories which have been put forward to explain the XCI phenomenon (Gartler and Riggs, 1983), one which has not been discounted holds that inactivation is a gene by gene event which is nucleated from one distinct

chromosomal site - the control element (Brown et al 1991) - and involves discrete cis-acting elements (Goldman et al 1987; Mohandas et al 1987; McBurney, 1988; Brown and Willard, 1990). No consensus sequence for a discrete cis-acting element is apparent (Goldman et al 1987), suggesting that either secondary structure is more important than sequence, or the discrete cis-acting regulatory elements lie outside the immediate promoter region. It is proposed by Goldman et al (1987) that the functional unit of the spread and maintenance of XCI is the chromatin loop domain, and that the signal sequences required for XCI occur only once per domain. Some domains would have this signal sequence and some would not, which would explain why some autosomal genes are inactivated in X-autosome translocations (the AFP minigene) (Krumlauf et al 1986) and others are not (the chicken transferrin gene) (Goldman et al 1987).

Both the human *PGK-1* and *HPRT* X-linked genes have been analyzed for binding of factors to their promoters. Using *in vivo* footprinting techniques, the promoters of these genes on the active and inactive X chromosomes were compared for binding activities. In both cases, the promoter of the allele on the inactive X chromosome was free of protein. On the *HPRT* promoter on the active X chromosome, six footprints were detected: four over GC-boxes, one associated with a potential AP-2 binding site, and one with a sequence not previously reported to interact with a sequence-specific DNA-binding factor (Hornstra and Yang, 1992). On the *PGK-1* promoter on the active X chromosome, 5-7 footprints were detected, depending on the footprinting method used. All but one of the footprints coincided with known consensus binding sites. There is no striking homology between the *PGK-1* and the *HPRT* footprinted regions which do not contain known consensus binding sites.

X chromosome inactivation also occurs in male germ cells. Presumably, however, the mechanism by which both the X and Y chromosomes become highly condensed and transcriptionally inert during spermatogenesis is different from the dosage compensation mechanism by which one of the two X chromosomes becomes transcriptionally inert during female development.

1.5 Transcriptional regulation of PGK

As mentioned above, binding studies examining the interaction of proteins with X-linked promoters identified multiple protein binding sites on the human *PGK-1* promoter. *In vivo* binding studies using both DMS and DNase 1 footprinting techniques (Pfeiffer et al 1990; Pfeiffer and Riggs, 1991) revealed binding to an "Nf-1 like" binding sequence, GC-boxes, a CCAAT-box and the consensus sequence for the transcription factor ATF (see Figure 5).

In vitro binding studies (Yang et al 1988) using a HeLa nuclear extract heparin-agarose 0.2 - O.4 M KCl fraction identified two protein binding sites by DNase 1 footprinting (at -360 and -130) and one by gel shift analysis (around -150).

The 5'-flanking sequences of the *pgk-1* promoter in the human and mouse are 61% homologous in the core promoter region (to -121), 79% homologous in the proximal region of the UAS (-121 to -213) and 66% homologous in the distal region of the UAS (-213 to -424). The mouse promoter contains four GC-boxes in the $Pgk\text{-}1^a$ allele (five in the $Pgk\text{-}1^b$ allele) and a CCAAT-box (see Figure 6A). A degenerate Nf-1 consensus site is present at the same site as the consensus Nf-1 site in the human

FIGURE 5: Consensus binding sites and protein:DNA interaction on the human *PGK-1* promoter.

The positions of the various consensus binding sites on the promoter are: **Ets:** (-433 to -430)*, (-360 to -357)*, (-355 to -352), (-253 to -250)*, (-231 to -228), (-225 to -222)*, (-189 to -186), (-63 to -60), (-14 to -11)*; **GC box:** (-245 to -240)*, (-58 to -53), (-34 to -29); **NF-1:** (-136 to -124); **CAAT box:** (-106 to -102) relative to the most 5' transcription start site. * signifies an inverted consensus sequence.

The *in vitro* DNase 1 footprinting of Yang et al (1988) produced two footprints (solid black boxes), (-374 to -351), (-145 to -118); the *in vivo* DNase 1 footprinting of Pfeiffer and Riggs (1991) resulted in five footprints and a hypersensitive site (arrows) (-365 to -335), (-254 to -234/-251 to -234), (-222 to -198/-218 to -186), (-160 to -122), (-112 to -87/-113 to -83), (-61 to -39/-62 to -32) and; the *in vivo* DMS footprinting of Pfeiffer et al (1990) produced seven footprints (-369 to -359), (-302 to -298), (-255 to -239), (-228 to -208), (-156 to -142/-157 to -137), (-108 to -102), (-65 to -37).

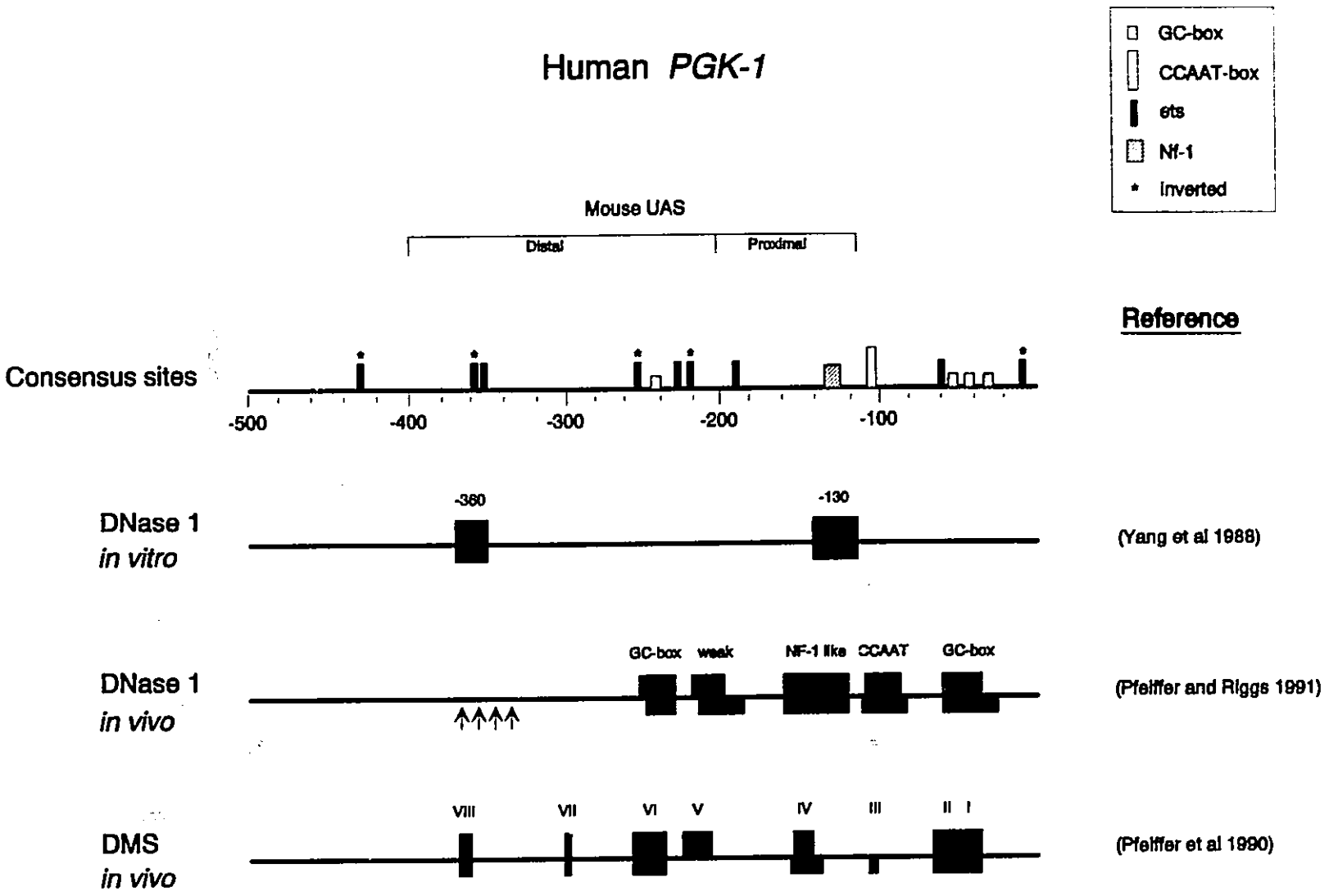
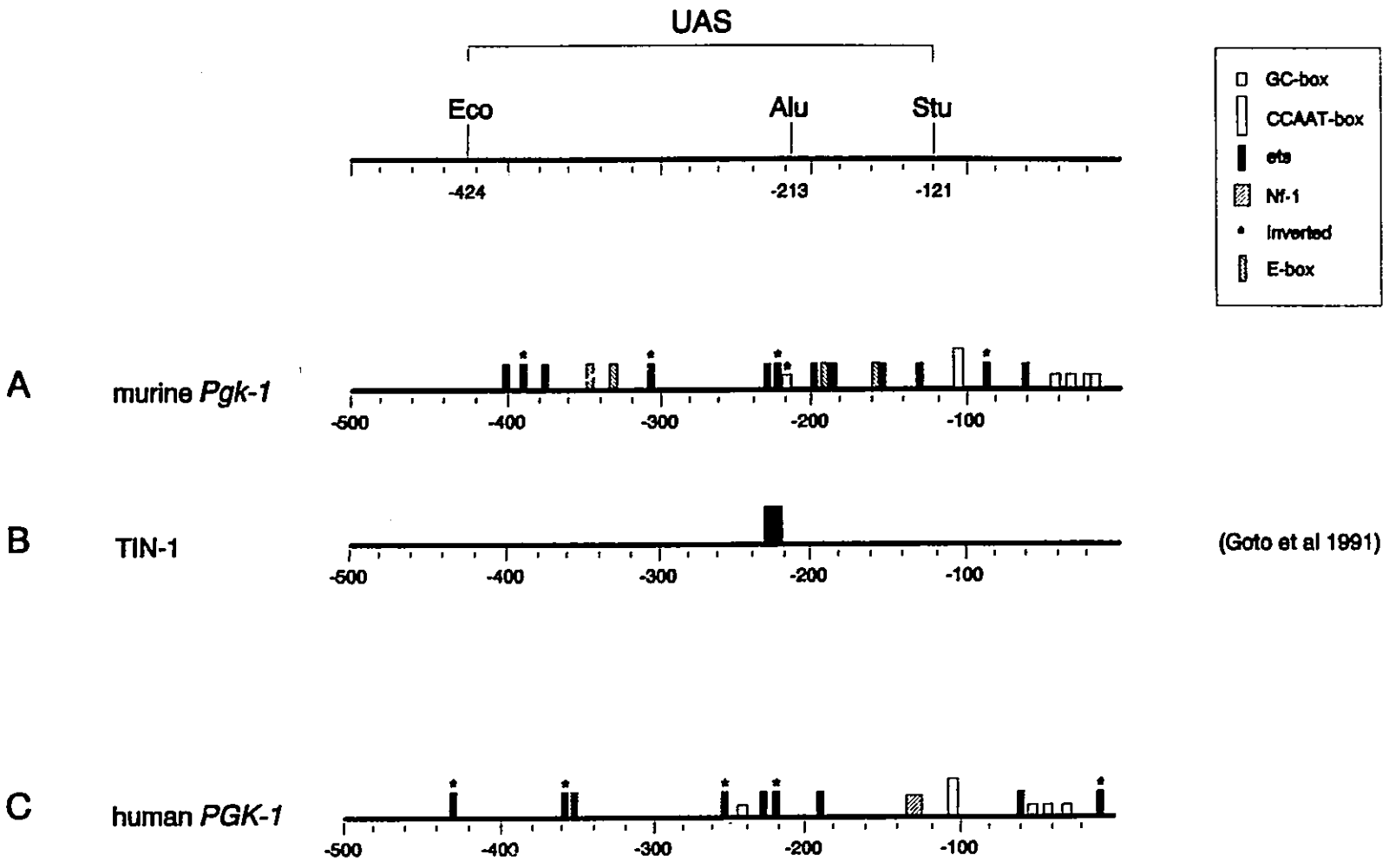

Human PGK-1
GC-box
CCAAT-box
ets
Nf-1
Inverted
Mouse UAS
Distal
Proximal
Reference
Consensus sites
-500
-400
-300
-200
-100
DNase 1
in vitro
-360
-130
(Yang et al 1988)
DNase 1
in vivo
GC-box
weak
NF-1 like
CCAAT
GC-box
(Pfeiffer and Riggs 1991)
DMS
in vivo
VIII
VII
VI
V
IV
III
II
I
(Pfeiffer et al 1990)

FIGURE 6: Consensus binding sites and protein:DNA interactions on the murine *Pgk-1* promoter.

(A) The positions of the various consensus binding sites on the promoter are: **Ets:** (-402 to -400), (-392 to -389)*, (-377 to -375), (-310 to -307)*, (-232 to -229), (-226 to -223)*, (-200 to -198), (-187 to -184), (-154 to -152), (-134 to -131), (-91 to -88)*, (-63 to -60); **GC box:** (-224 to -219)[1]*, (-46 to -41), (-34 to -29), (-22 to -17), (-18 to -13); **E-box:** (-349 to -344), (-335 to -330), (-192 to -187), (-159 to -154); **CAAT box:** (-106 to -102) based on the sequence in Boer et al (1990) and relative to the most 5' transcription start site, outlined in McBurney et al (1991). * indicates inverse sequence, [1] indicates sequence present in pgk-1^b only.

(B) Factor interaction with the promoter. TIN-1 is the testis-inhibitor factor from rat liver nuclear extract (Goto et al 1991) which recognizes the palindromic Ets site between (-232 and -223).

(C) The consensus binding sites of the human Pgk-1 promoter are included for comparison purposes.


UAS
Eco
Alu
Stu
-424
-213
-121
GC-box
CCAAT-box
ets
Nf-1
Inverted
E-box
A
murine Pgk-1
-500
-400
-300
-200
-100
B
TIN-1
-500
-400
-300
-200
-100
(Goto et al 1991)
C
human PGK-1
-500
-400
-300
-200
-100

promoter. There is no ATF binding site. Ets consensus sequences are present throughout the promoter.

Two transcription factors have been identified which bind to the promoter of *pgk-1*, RAP1 (see below) and TIN-1.

TIN-1 is a 43-45 kD testis-specific factor from rat (Goto et al 1991). It binds to the sequence 5'-GGAATTTCC-3' which contains a palindrome of a normal and inverted Ets consensus binding sequence. The same sequence is present in the mouse promoter, with a slight variation between the *Pgk-1*^{a} and *Pgk-1*^{b} alleles (the *Pgk-1*^{a} allele has the sequence 5'-GGAA(G)TTCCC(CC)CCGC-3', and the *Pgk-1*^{b} allele has the sequence 5'-GGAA(GT)TTCCCCC(A)GC-3') (see Figure 6B). TIN-1 has been shown to inhibit transcription of the *PGK-1* gene (Goto et al 1991) but activate transcription of the tissue-specific *PGK-2* gene during spermatogenesis (Goto et al 1993). This demonstrates the importance of the Ets binding motif in phosphoglycerate kinase regulation.

In *Pgk-2*, the region to -190 serves as a core promoter, and an additional 327 base-pairs is necessary for the testis-specific expression of the gene in transgenic mice (Robinson et al 1989) (see Figure 7). Therefore, the expression of both PGK-1 and PGK-2 is modulated by an enhancer or activator sequence 5' to the core promoter.

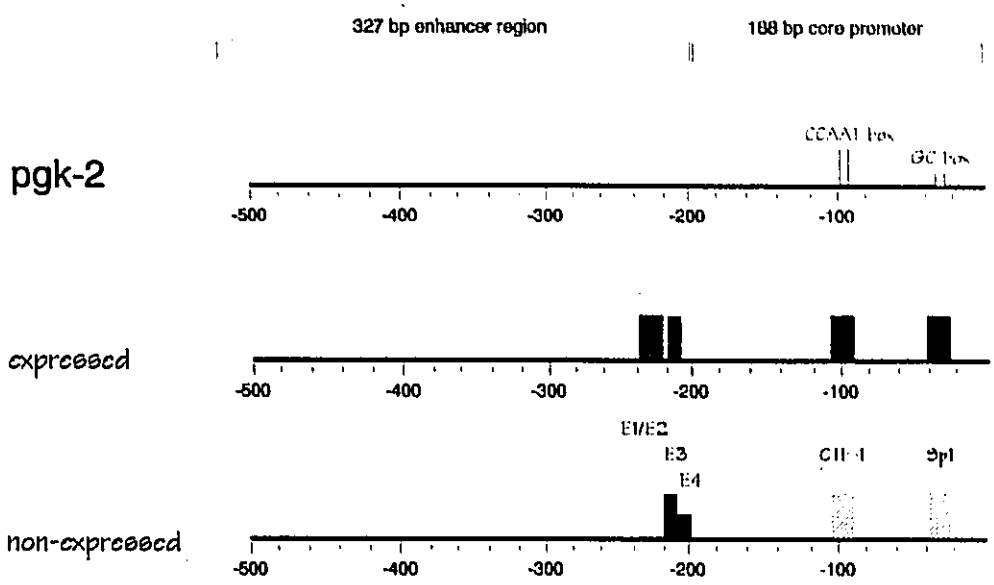
RAP1 is a yeast transcription factor which binds to the yeast *Pgk-1* promoter. Carbon source regulation of phosphoglycerate kinase transcription was found to be mediated, in part, through regulation of this 93 kD RAP1 protein (Shore and Nasmyth, 1987). RAP1 binds to the activator core (AC) sequence (Stanway et al 1987; Chambers et al 1989), as illustrated in Figure 8. The AC sequence (5'-AAACCCAGACAC-3') lies

FIGURE 7: Protein:DNA interaction associated with the expressed and non-expressed state of the murine *Pgk-2* gene.

(A) In the expressed state, the general transcription factors Sp1 and CTF-1 bind to the core promoter, in addition to two proteins (designated E_3 and E_1/E_2) which bind just 5' to the core promoter in the region necessary for testis-specific expression. E_3 is a ubiquitous factor, and also binds to the tissue-specific region of *Pgk-2* in the non-expressed state. E_4 also binds in the non-expressed state. It is unknown whether CTF-1 and Sp1 bind to the core promoter of the non-expressed *Pgk-2* gene (Gebara and McCarrey, 1992). The DNA binding activities were detected by gel shift analysis, comparing nuclear extracts from expressing adult mouse testis cells to non-expressing prepuberal testis cells and non-expressing somatic cells.

(B) Delineation of the sequences of the E_1/E_2 and E_3/E_4 regions.

A



B

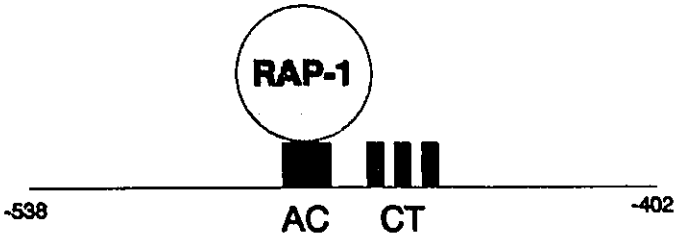
E1/E2 region : 5'-CCAGCAAAAAAGTTCA-3'

E3/E4 region : 5'-CCAAATCAAGATAGACAG-3'

FIGURE 8: Functional elements within the yeast *PGK* upstream activator sequence.

(A) The upstream activator sequence, between -538 and -402 relative to the translation start site, contains the activator core sequence (AC), and three copies of the pentamer sequence (CT). The activator core binds the transcription factor RAP-1. **(B)** Delineation of the sequences of the activator core and the pentamer block.

A



B

AC sequence/RAP-1 binding site

5'-AAACCCAGACAC-3'

Pentamer sequence

5'-CTTCC-3'

within the UAS of the yeast *Pgk* gene, between bases -538 and -402 relative to the initiating ATG (Ogden et al 1986). The UAS contains multiple functional elements, including the essential AC sequence and three copies of the pentamer sequence 5'-CTTCC-3' (Stanway et al 1987; Chambers et al 1988). The promoters of many other yeast glycolytic genes also contain RAP1 binding sites and copies of the pentamer sequence, which suggests a common or coordinated pathway for the expression of the glycolytic genes in yeast.

The consensus binding sites within the promoter of the murine and human *pgk-1* gene are for the Ets family of transcription factors, bHLH transcription factors, GC-box binding factors, NF-1 family factors, CCAAT-box binding factors, the ATF transcription factor in the human *PGK-1* promoter, and the RAP1 protein in the yeast promoter.

GC box binding factors include Sp1, ETF, and the repressor protein GCF (Kageyama et al 1989).

The Nf-1 family of transcription factors includes NF-1, CTF-1 and CTF-2 (Santoro et al 1988; Ristiniemi and Oikarinen, 1989). NF-1 potentially binds to the palindromic sequence 5'-TGGA/C(N)$_5$GCCA-3' (Gronostajski, 1987; Chodosh et al 1988; Nilsson et al 1989), but can also bind to the TGGCA "half-site" (Gil et al 1988). CTF-1 and CTF-2 have a preference for the sequence 5'-A/GA/GCCAAT-3' (Chodosh et al 1988). Each protein is apparently composed of at least two different subunits (Chodosh et al 1988). The proteins bind to either half sites or full sites as dimers (Gounari et al 1990), but highest affinity binding occurs as heterodimers, for example CTF-1A and CTF-1B (Chodosh et al 1988).

NF-1 functions in both transcription and replication and is present in a wide variety of cells. However, NF-1 family members are apparently not present in high concentrations, if at all, in undifferentiated cells. Binding activity of NF-1 was not detected in the mouse erythroleukemia line, MEL, nor the embryonal carcinoma (EC) cell lines F9 or PCC4 (Speck and Baltimore, 1987; McQuillan et al 1991). Goyal et al (1990) found only 10% binding activity of NF-1 in embryonic stem (ES) cells and F9 cells compared to HeLa cells. In P19 EC cells, NF-1 is unable to bind to the human papovavirus JC viral promoter, whereas in P19 cells induced to differentiate with retinoic acid, all three NF-1 binding sites are protected from DNase 1 digestion (Nakshatri et al 1990).

The bHLH family of transcription factors binds the E-box consensus sequence 5'-CANNTG-3'. In this sequence, the central two nucleotides are usually C and G (Blackwell and Weintraub, 1990). For instance, c-Myc and USF bind the sequence 5'-CA(CG)TG-3', AP-4 binds the sequence 5'-CA(GC)TG-3', and CBF-1 binds the sequence 5'-CA(CA/G)TG-3'. With so few differences in the core sequence, Blackwell and Weintraub (1990) suggest that the E-box is a half-site recognition sequence, where bordering nucleotides are important in binding site selection.

The Ets family of proteins has been implicated in regulation of transcription, cell transformation and development (Seth et al 1989; Seth and Papas, 1990; Wasylyk et al 1992). Ets1 and Ets2 are synthesized when mouse fibroblasts are stimulated by serum (Bhat et al 1987) and are phosphorylated in response to mitogenic signals (Fujiwara et al 1990; Pognonec et al 1990), suggesting that at least the activities of Ets1 and Ets2 are

regulated by both pre- and post- translational mechanisms (Wasylyk et al 1992). Phosphorylation had no effect on binding activity, and it is not known at this time how it affects transcriptional activation (Wasylyk et al 1992).

Expression of Ets1 and Ets2 was examined in human resting T-cells, and compared to expression in T-cells activated either by cross-linking of the T-cell receptor-CD3 complex on the cell surface, or by direct stimulation with phorbol esters and ionomycin (Bhat et al 1990). Resting cells only expressed Ets1. Upon activation, Ets1 expression decreased until its mRNA was undetectable, then was reinduced and maintained at a high level. Ets2 expression was undetectable in resting T-cells, present in activated cells, then decreased again to undetectable levels. Ets1 and Ets2 therefore have a reciprocal relationship in the regulation of human T-cell activation; Ets1 may be necessary for maintenance of the quiescent state, whereas Ets2 is associated with cellular activation and proliferation (Bhat et al 1990).

According to a recent report by Kola et al (1993), Ets2 is expressed ubiquitously, whereas Ets1 is present only in mesodermal cells associated with morphogenetic processes such as organ formation and tissue modelling. Ets2 was expressed in differentiated and undifferentiated P19 and F9 cells, and in all organs of embryonic, neonatal and adult mice studied. Ets1 was undetectable in either differentiated or undifferentiated F9 cells (Martin et al 1992), but detectable in P19 cells induced to differentiate into mesoderm-like cells. It was suggested by the authors that Ets1 has an important role in morphogenetic processes whereas Ets2 plays a more fundamental role in cells (Kola et al 1993).

In summary then, Ets-1 is upregulated in the resting T-cell and the differentiated P19 cell, but downregulated in the active T-cell or the very actively cycling undifferentiated P19 cell.

In mouse, mRNA from the Ets-family factor PEA3 is detectable only in epididymis and brain, and in F9 and P19 EC cells prior to differentiation by retinoic acid to extraembryonal endoderm and neurons, respectively (Xin et al 1992; Martin et al 1992); Spi-1 is expressed preferentially in B cells and macrophages (Klemsz et al 1990), and; Fli-1 is expressed in erythroleukemic cells as well as in the thymus and spleen (Ben-David et al 1991).

Both Ets1 and Ets2 share many properties with Jun, Fos and Myc: nuclear localization, rapid turnover, and quick response to second messengers (Bhat et al 1990). In fact, Ets1 cooperates with the c-Fos and c-Jun transcription factors in transcriptional activation of the polyoma enhancer (Wasylyk et al 1990). Ets2 is also believed to function in concert with another factor since Ets2 alone will not transactivate. On the mim-1 promoter, Ets2 alone fails to transactivate, and Myb (a bHLH transcription factor) alone leads to low levels of transcription; however, the presence of both Ets2 and Myb results in elevated transcription (Dudek et al 1992). Ets1 failed to affect trans-activation. As both Myb and Ets2 bind to the mim-1 promoter, and no physical association has been detected by immunoprecipitation, it has been suggested that the two proteins' cooperativity in transcriptional activation may not be due to direct binding between the proteins and may involve coactivator molecules.

1.6 P19 cells

P19 embryonal carcinoma (EC) cells were isolated from a teratocarcinoma produced by transplanting a 7.5 day C3H/He strain mouse embryo into the testis of another adult mouse. Stem cells of the transplanted embryo were isolated from the primary tumor to form the immortalized P19 EC cell line (McBurney and Rogers, 1982). P19 cells have a euploid male karyotype, and are maintained in an undifferentiated state by routine subculturing in serum-supplemented media.

P19 cells resemble the pluripotent stem cells of the early embryo in their capacity to differentiate. If placed back into a blastocyst, which is subsequently implanted into a pregnant mouse, the resulting P19-derived cells are present in a variety of apparently normal tissues, although most of the embryos with large contributions of P19 cells are abnormal in some way (Rossant and McBurney, 1982).

The differentiation of these cells can be controlled by drugs at non-toxic concentrations. Treating cells in culture with retinoic acid (RA) (10^{-7}M) induces the formation of glial cells, neurons and astrocytes - cell types normally derived from the neurectoderm (Jones-Villeneuve et al 1982). Treatment of the cells with dimethyl sulfoxide (DMSO) (0.5 - 1.0 %) induces the formation endodermal and mesodermal derivatives, including cardiac and skeletal muscle cells (McBurney et al 1982). As P19 cells can also be effectively transfected with recombinant DNA, they provide a convenient system for investigating the molecular mechanisms governing developmental decisions made by differentiating pluripotent cells.

1.7 Project objectives

The objective of my research was to investigate the mechanism of transcription factor regulation of the murine *Pgk-1*^{b} gene.

My hypothesis was that PGK-1 expression is down regulated in differentiated P19 cells, which are not as metabolically active as undifferentiated, actively cycling P19 cells. This developmental and tissue-specific regulation of gene expression is controlled by transcription factors which bind elements within the promoter of the *Pgk-1* gene. The same factors, or subset of factors, which interact with the promoter in cells where *pgk-1* is transcriptionally active, are also involved in maintaining the gene in a transcriptionally active configuration on the active X chromosome in female cells.

The purpose of my thesis project was specifically to identify cis-acting DNA binding elements within the upstream activator sequence of the *Pgk-1* gene, and to demonstrate their involvement in the developmental transcriptional regulation of *Pgk-1* gene expression.

Towards this end, my research began with an investigation of protein-DNA binding to the murine *Pgk-1* upstream activator sequence, including a comparison of binding activity in this region between male and female cell lines, bearing in mind the potential involvement of trans-acting factors in X chromosome inactivation. The results of these studies are presented in Chapter 3.

The effect of differentiation on *Pgk-1* gene expression was explored using P19 cells induced to differentiate with retinoic acid. Expression levels before and after differentiation were correlated with binding activity in the upstream activator sequence.

The results of these experiments are presented in Chapter 4.

Finally, in Chapter 5, a model is presented which suggests a mechanism to explain the observed correlation between protein binding to the upstream activator sequence and gene activity.

CHAPTER 2

Materials and Methods

2.1 Culture Conditions for EC Cells

P19 EC cells were maintained in culture and induced to differentiate as described by Rudnicki and McBurney (1987).

2.1.1 Routine Culture

EC cells were cultured in a 5% CO_2 incubator at 37°C in the presence of α-modified minimal essential medium (Gibco, Mississauga, Canada) supplemented with 7.5% calf serum and 2.5% fetal calf serum (Bocknek Laboratories, Rexdale, Ontario, Canada or FetalClone II, HyClone, Logan, Utah, USA). The cells were passaged every two days. Briefly, cells were washed with PBS (0.8% NaCl, 0.02% KCL, 0.02% KH_2PO_4, and 0.115% Na_2HPO_4, pH 7.4), and were incubated for 5 min with a small volume (0.5 ml

in a 60 mm dish) of trypsin-EDTA (1 mM EDTA and 0.025% w/v trypsin in PBS). The cells were dispersed by vigorously pipetting with a Pasteur pipette, and were resuspended in medium before being transferred to tissue culture dishes at a density of approximately 10^5 cells/ml of medium. If necessary, cells were counted with the aid of an electronic counter (Coulter Counter, Coulter Electronics Inc.).

2.1.2 Conditions for Differentiation

P19 cells were induced to differentiate into neurons, glia and fibroblasts by culturing in tissue culture dishes in the presence of 0.3 μM retinoic acid (RA) (Eastman Kodak). The medium was replenished every two days. For cultures exposed to RA for eight to ten days, the cells were passaged every two to three days prior to the appearance of neurons so the dishes would not be confluent for transient transfections.

2.2 Transfection and Selection of EC Cells

Transfections of EC cells with plasmid DNA were performed according to the protocol of Chen and Okayama (1987). Highly purified plasmid DNA for transfections was prepared as described in Section 2.6.2. The day preceding the transfection, P19 cells were seeded at a density of 500,000 per 60 mm dish. The following day, the medium (5 ml) was changed about one hour before addition of the DNA to the dishes. Plasmid DNA (10-20 μg) was added to 0.5 M $CaCl_2$ (500 μl), followed by 500 μl of 2X BES buffer (50 mM N,N-bis 2-hydroxyethyl-2-aminoethane-sulfonic acid pH 6.90, 280 mM NaCl, 1.5 mM Na_2HPO_4). The solution was vortexed and stored 5 to 10 min at room

temperature. The precipitated DNA was added dropwise to the cells and remained in the dish for 6 to 8 hours. The medium was then changed, and the cells grown an additional 24 hours. Next, the cells were either harvested, if a transient assay were being done, or plated in the presence of the antibiotic puromycin, if clones were selected. For the selection of transfected clones, 500,000 cells were plated onto a 100 mm dish with 10 ml of medium, or 2 x 10^6 cells were plated onto a 150 mm dish with 25 ml of medium. The selective medium for the puromycin gene product consisted of regular medium supplemented with 2 μg/ml puromycin. The medium was usually changed 3 days after the beginning of selection. It was also changed later if a large number of dead cells was present in the medium during the 8 days of selection. All transfection experiments were performed a minimum of two times. To ensure that there was efficient transfection, an internal standard was included along with the test plasmid.

β-galactosidase activity was measured using the spectrophotometric assay developed by Norton and Coffin (1985). After three washes in PBS, cells were harvested in a sucrose solution (0.25 M sucrose, 10 mM Tris-HCl, 10 mM EDTA) by scraping. Cells were lysed by three freeze/thaw cycles, and the debris was pelleted by centrifugation for 10 min in a microcentrifuge. Total protein concentration in each cell extract was determined using the Bio-Rad protein assay solution, based on the Bradford method (Bradford, 1976). The reaction mixture for the β-galactosidase assay consisted of 800 μl of assay buffer (100 mM $NaPO_4$, pH 7.0, 1 mM $MgSO_4$, 100 mM β-mercaptoethanol), 200 μl of stock ONPG (4 mg/ml ONPG in assay buffer), and 400 μg of protein from the cell extracts. Final volumes were equalized by addition of sucrose

solution. The reaction at 28°C was monitored by measuring the change in absorbance at a wavelength of 420 nm over 30 min. One unit of β-galactosidase activity is the activity that permits the hydrolysis of 10^{-9} moles of ONPG/min in the assay conditions described.

Chloramphenicol acetyltransferase (CAT) activity was assayed by measuring the acetylation of chloramphenicol by ^{3}H acetyl coenzyme A in a modified version of the Sleigh protocol (1986). Cell extracts were prepared in a sucrose solution as described for the β-galactosidase assay. Total protein concentration in the cell extract was determined with the Bio-Rad protein assay solution. Cell extracts (200 μg of protein) were first heated at 65°C for 10 min. Following centrifugation, the supernatant was incubated for one hour at 37°C in the presence of 1.6 mM chloramphenicol, 0.4 μCi of ^{3}H acetyl CoA (4.3 Ci/mmol), 0.25 M Tris-Cl, pH 7.8 in a total volume of 200 μl. The reaction mixture was extracted twice with 400 μl of cold ethyl acetate, and the amount of ^{3}H in the supernatant was measured in a liquid scintillation counter.

2.3 Extracts

Nuclear extracts were made from P19 cells and used in DNA binding studies. Initially, binding studies were conducted using total nuclear extract, obtained by either of two modified methods of Dignan et al (1983). Later, however, binding studies involved the use of nuclear extract fractionated over either heparin-agarose or sephacryl columns.

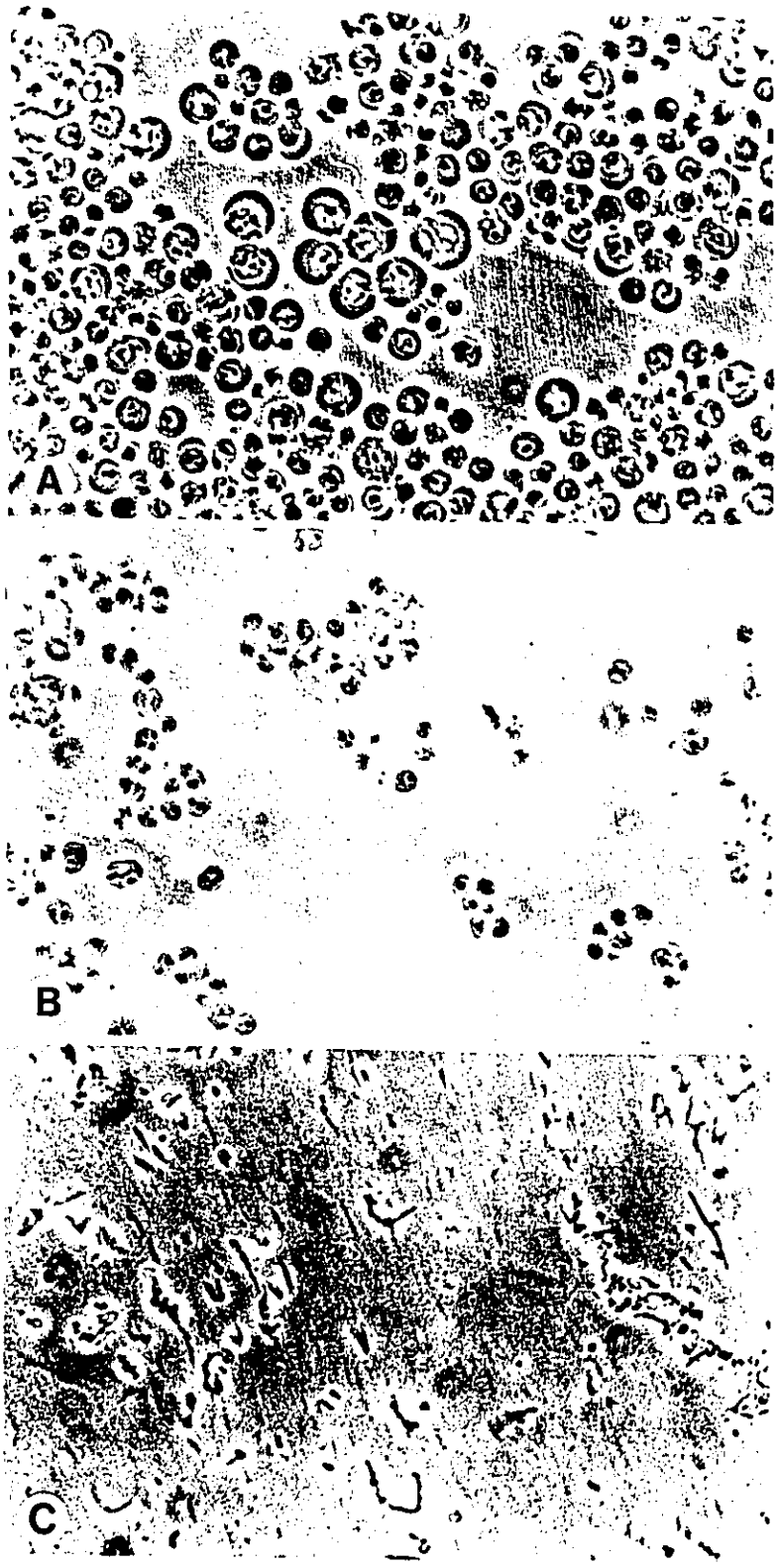
2.3.1 Total Nuclear Extract

Monolayer P19 cells at approximately 80% confluence were harvested in a solution of 1 mM EDTA and 0.025% trypsin in PBS. Cells were washed in PBS, centrifuged at 800 g for 3 min (in a Jouen CR4.11 centrifuge), resuspended in 5 times the packed cell volume of Buffer A (10 mM Tris-HCl, pH 8.0, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT) and left to swell on ice for 10 min (Figure 9A). They were then centrifuged at 400 g for 10 min and the pellet resuspended in 2 times the packed cell volume of fresh Buffer A. To lyse the cells 0.25 M sucrose and 0.1% Nonidet P-40 were added, followed by 0.8 M sucrose to stabilize the extract (Figure 9B). The mixture was centrifuged at 1000 g for 10 min. After removal of the supernatant (the nuclear-free cell extract) the nuclei were washed by resuspension in at least two volumes of Buffer A and centrifuged (1000 g for 10 min). The crude nuclear pellet was resuspended in 3 ml/10^9 cells of Buffer C, followed by incubation on ice for 30 min with occasional vortexing. For maximum yield of nuclear material, the mixture was forced through a 20 gauge needle about 5 times (Figure 9C). The nuclear debris was then pelleted by centrifugation for 20 min in a microcentrifuge (12,000 g). The supernatant was transferred to a new tube and mixed with 0.33 g/ml of NH_4SO_4. After 30-60 min of rotation at 4°C the nuclear extract was centrifuged at 12,000 g for 20 min. The pellet was resuspended in approximately one half volume of Buffer D and dialysed, along with the nuclear-free cell extract, against 100 volumes of the same buffer for 3-6 hours.

To confirm that nuclear and nuclear-free cell extracts were free from cross-contamination, both were assayed in a band shift assay using the AdMLP (adenovirus

FIGURE 9: Microscopic view of P19 cells during nuclear extract preparation.

(A) Swollen cells, after being resuspended in Buffer A and maintained on ice for 10 min. **(B)** Intact nuclei, after cell membrane lysis with detergent. **(C)** Debris, following nuclear disruption by passage through a syringe. Photographs were taken with Kodak TMAX 100 film. Bar, 20 μm.


A
B
C

major late promoter). A 130 bp Hind III-Pvu II restriction fragment of AdMLP DNA was used as template. Three bands, representing complexes of AdMLP DNA-binding proteins present in the nuclear extract but not the nuclear-free cell extract, were indicative of adequate separation of extracts for binding analyses (Carthew et al 1985) (Figure 10).

To ensure that protein was not degraded, samples were electrophoresed by SDS-PAGE (see Figure 11).

Nuclear extract from RA-treated P19 cells was compared to nuclear extract from untreated P19 cells by band shift analysis (see below) to demonstrate the existence of a different population of factors. The WiF-1 double stranded oligonucleotide (see below) was used as template. In RA-treated P19 cells a new binding activity is observed which is absent in P19 nuclear extracts (see Figure 12) (St-Arnaud and Moir, 1993).

2.3.2 Fractionated Nuclear Extract

P19 nuclear extract was fractionated using a heparin-agarose column (Sigma Heparin-agarose Type II in a Bio-Rad Econo-Column 10). All buffers contained 50 mM Tris-HCl, pH 7.4, 50 mM $MgCl_2$, 5 mM DTT, 20% glycerol and variable amounts of KCl. Columns were prepared and washed by placing 2 ml of gel solution in a 5 ml polystyrene tube, letting it settle, and aspirating off the liquid fraction. 1 ml of buffer containing 1.0 M KCl was added to the resin, which was vortexed. After settling, the liquid fraction was aspirated. This step was repeated. 1 ml of buffer containing 0.1 M KCl was added, the resin stirred, and placed in the column to which two drops of buffer had already been

FIGURE 10: Separation of nuclear and nuclear-free cell extracts.

The adenovirus major late promoter 130 bp Hind III-Pvu II fragment, end-labelled with (γ-^{32}P)ATP and incubated with either nuclear extract (NE) (lanes 1-4) or nuclear-free cytoplasmic extract (NFCE) (lanes 5-8) from P19 cells, at the protein concentrations indicated. Complexes III, II and I appea. as in Carthew et al (1985).

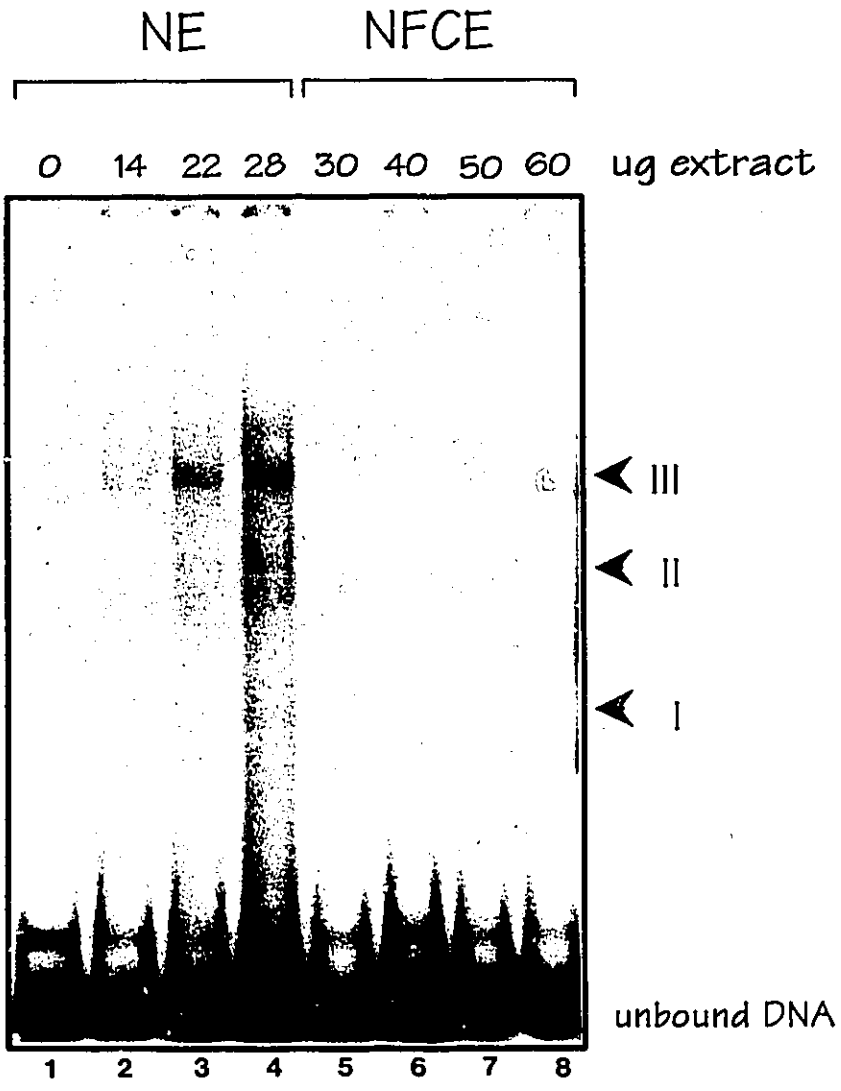

NE
NFCE
0 14 22 28 30 40 50 60 ug extract
III
II
I
unbound DNA
1 2 3 4 5 6 7 8

FIGURE 11: P19 nuclear extract visualized by SDS-PAGE.

A 7.5% denaturing gel was used. Lanes 1 and 2: 27 μg and 21 μg, respectively, of two independent preparations of P19 nuclear extract; lane 3: 9 μg of Sigma-7-B pre-stained molecular weight marker. The gel was electrophoresed at 200 volts for 45 minutes, and stained with Coomassie Brilliant Blue R250.

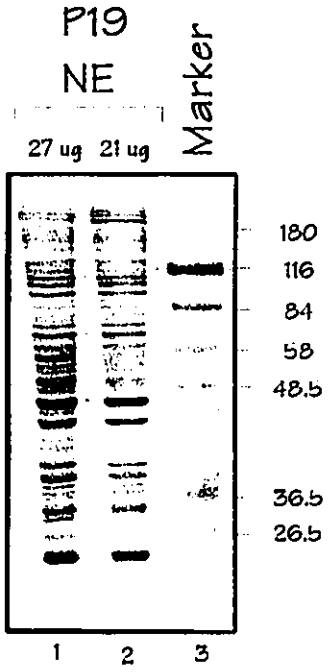
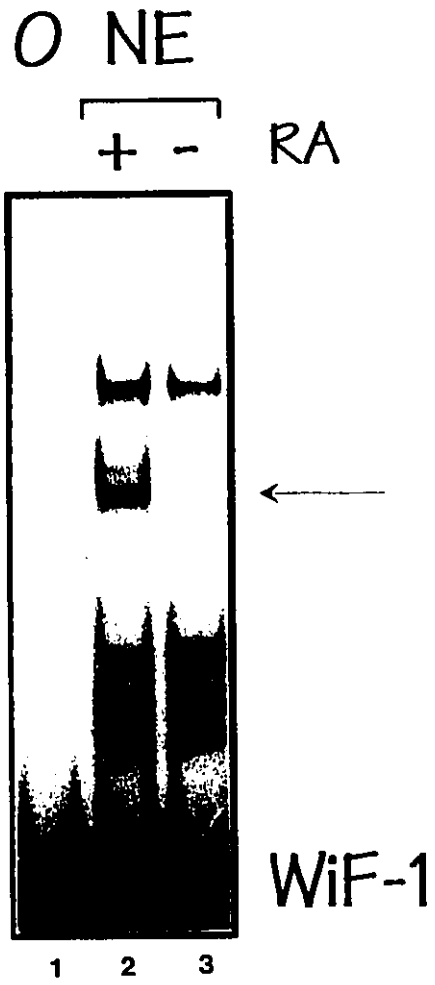

P19
NE
27 ug
21 ug
Marker
180
116
84
58
48.5
36.5
26.5
1
2
3

FIGURE 12: RA-treated cells have a different population of DNA binding factors.

Band shift of the WiF-1 oligonucleotide incubated with untreated or RA-treated P19 cell nuclear extract. 0.5 ng of DNA were mixed with 20 μg of nuclear extract and 3 μg of poly(dI-dC).


O NE
+ -
RA
WiF-1
1 2 3

added. After the resin was left to settle, it was drained through the column. 2 ml of buffer containing 0.1 M KCl was followed by 2 ml of buffer containing 1.0 M KCl followed by 4 ml of buffer containing 0.1 M KCl. At this point P19 nuclear extract was added.

500 μl of extract from five different nuclear extract preparations were used, roughly 6 mg of protein. 500 μl aliquots were collected in 17 tubes. 2 ml of buffer containing 0.1 M KCl were passed over the column, followed by 2 ml aliquots of buffer containing 0.24 M KCl, 0.6 M and 1.0 M KCl. Eluted fractions were dialysed overnight against 500 ml of Buffer DE-0.5 (10 mM Tris-HCl, pH 7.4, 50 mM KCl, 5 mM $MgCl_2$, 0.5 mM DTT and 20% glycerol). Protein concentrations were determined using the Bradford assay. Fractions 3, 7 and 11 had peak concentrations. Fractions were next concentrated in Centricon-10 micro-concentrators. 500 μl fractions were added to the concentrator and centrifuged at 5000 g for one hour in a Beckman JA-17 rotor after which time the volume was reduced to approximately 200 μl. 10 μl of protein were electrophoresed on a 10% SDS polyacrylamide gel to monitor degradation and DNA contamination. The fractions were tested for DNA binding activity in band shift assays.

Nuclear fractions from P19 cells exposed to RA for four days were the kind gift of Dr. St-Arnaud (Shriners Hospital for Crippled Children and McGill University, Montreal), and were prepared as described (St-Arnaud and Moir, 1993). For the "NE-17" preparation, crude nuclear extract was chromatographed over a wheat germ agglutinin column, precipitated in 53% ammonium sulfate, and further chromatographed over a Sephacryl-300 column. The "NE-23" and "NE-24" preparations were simply

chromatographed over Sephacryl-300. I was given fractions 10-30 of NE-17, 10-15 and 28-40 of NE-23, and 16-19 and 25-35 of NE-24. The middle fractions of the NE-23 and NE-24 preparations were retained by Dr. St-Arnaud as they contained the WiF-1 protein he was purifying.

2.4 DNA Templates

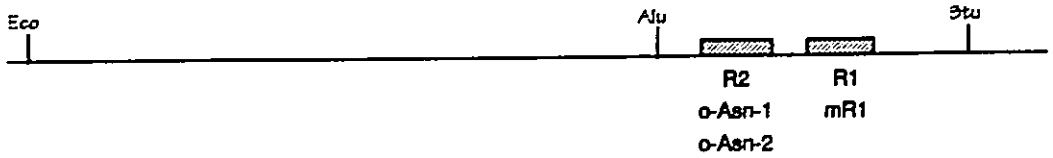
2.4.1 Oligonucleotides

Five double-stranded oligonucleotides were used as templates in the various DNA binding experiments (Table 1). Two single-stranded oligonucleotides encompassing the WiF-1 binding site of the Wnt-1 gene were a gift from Dr. St-Arnaud. The sequence for the Nf-1 double-stranded oligonucleotide used in the determination of Nf-1 presence in P19 nuclear extracts was obtained from a paper by Gronostajski (1987). A double-stranded oligonucleotide with the sequence of the R1 region, and another incorporating a mutation in the same region, were used. The R2 sequence was also incorporated into a double-stranded oligonucleotide. Two single-stranded oligonucleotides were designed for use in the PCR reaction to create a mutation of the R2 region (0-Asn-1, 0-Asn-2). Oligonucleotides were generated on either an Applied Biosystems 380B or 391 DNA synthesizer. To anneal the single-stranded oligonucleotides for DNA-binding studies, they were heated at 65°C for 15 min, 37°C for 30 min and room temperature for one hour.

TABLE 1: Nucleotide sequence of oligonucleotides

Designation	Nucleotide sequence	Ends generated
WIF-1	5'-GGGGGCGGGGGTGAGCCCCGA-3' *	-
Nf-1	5'-CGGAGGTCTGGCTTTAAGCCAAGAGCCGCC-3' *	Hpa II
mR1	5'-AGCTTCTCGTGGTGATGGACAGCACCGCTGAGCAAA-3' *	Hind III
R1	5'-CTCGTGCAGATGGACAGCACCGCTGAGCAA-3' *	-
R2	5'-GATCTGGACGTGACAAATGGAAGTAGCACGTCA-3' *	Bgl II
o-Asn-1	5'-AAATAATTAATTTCACGTCCTGCACGACGCGAGC-3'	-
o-Asn-2	5'-AAAAAATTAATTAGTAGCACGTCTCACTAGTCTC-3'	-

* The sequence of the upper DNA strand of each double-stranded oligonucleotide is shown



2.4.2 Activator Fragments

Plasmid pCAB17prX/T520, constructed by C.Adra, contains 500 bp of the *Pgk-1* 5'-flanking region in pGEM4. Plasmid pLSe140 was created by subcloning the Hha-Stu *Pgk-1* fragment from pCAB17prX/T520 into the Sma 1 polylinker site of pTZ19R, having first blunt-ended the Hha 1 end. Two plasmids were obtained, one having one insert, the other containing two inserts in a head-to-tail arrangement. DNA fragments used in binding assays are outlined in Table 2.

2.4.3 Transfected DNA

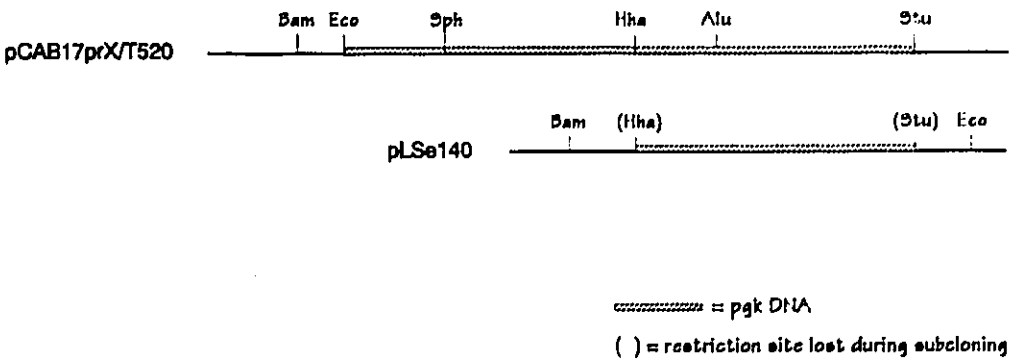
pET.lacZ contains 530 bp of the *Pgk-1* promoter between the Eco R1 and Taq 1 restriction sites, in pGEM4, whereas pST.lacZ only contains 220 bp between the Stu 1 and Taq 1 sites (McBurney et al 1991). pRSV.lacZ contains 524 bp of the RSV promoter driving lacZ. pKJ7.lacZ, constructed by K.Jardine, contains 500bp of the *Pgk-1* promoter, including the first three exons. pKJ17.CAT, also constructed by K.Jardine, contains 500 bp of the *Pgk-1* promoter (Xba 1-Taq 1), 730 bp of the CAT gene (Ban 1-Hind III) and 300 bp of the 3'-untranslated region of the *Pgk-1* gene (Bam H1-Hind III). The *Pgk*-puromycin plasmid contains 500 bp of the *Pgk-1* promoter driving the gene which encodes the antibiotic puromycin. B17 is a plasmid which contains a 17 Kbp Bam H1 fragment including the *Pgk-1* promoter and all of the *Pgk-1* gene up to and including exon 8 (Adra et al 1988). pKJ17mR2.CAT was constructed in Section 2.5 by mutagenesis.

TABLE 2:

Designation and description of DNA fragments routinely used in binding studies.

Fragment Designation	Length Overall (bp)	Length pgk	Plasmid	Digestion
SS	251	251	pCAB17prX/T520	Sph 1 - Stu 1
BS	326	320	pCAB17prX/T520	Bam H1 - Stu 1
BE	158	136	pLSe140	Bam H1 - Eco R1
ES	303	303	pCAB17prX/T520	Eco R1 - Stu 1
AS	92	92	1	Alu 1 - Stu 1
EA	211	211	1	Eco R1 - Alu 1

1 = purified from previously isolated BS fragment



2.4.4 Labelling

DNA fragments used in binding experiments were end-labelled using T4 polynucleotide kinase and (γ-^{32}P)ATP. All labelled fragments were purified from non-denaturing 8% polyacrylamide gels by elution at 37°C overnight in buffer containing 0.5 M NH_4OAc, 10 mM $Mg(OAc)_2$, 1 mM EDTA, 0.1% SDS.

2.5 Mutagenesis

pKJ17 was incubated with the o-Asn-1 and o-Asn-2 oligonucleotides (Table 1) in a reaction which included 1X vent buffer (10 mM KCl, 20 mM Tris-HCl (pH8.8), 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100), 100 μg/ml BSA, 4 mM $MgSO_4$, 200 mM each dNTP, 10 ng of pKJ17, 50 pmols of both oligonucleotides and 8 units of Vent(exo-) enzyme (New England Biolabs), in a total reaction volume of 100 μl. The reaction proceeded at 97°C for 1 minute, 66°C for 1 minute, 75°C for 6 minutes and was repeated during 28 cycles in a COY Tempcycler. The DNA was extracted with phenol and chloroform, precipitated with 0.3M sodium acetate and ethanol, and rinsed in 70% ethanol. The DNA was then treated with proteinase K (Crowe et al 1991). The dried pellet was resuspended in a buffer consisting of 10 mM Tris-HCl (pH8.0), 5 mM EDTA and 0.5% SDS. 1 μl of a 5 μg/μl stock of proteinase K (BRL) was added to the DNA which was incubated at 37°C for 30 minutes, then 68°C for 10 minutes. The DNA was extracted with phenol and chloroform, precipitated with 0.3 M sodium acetate in ethanol and resuspended in 50 μl of water. The DNA was then blunt ended, gene cleaned (Bio/Can Scientific) and ligated. Ligations were performed in 30 μl using 20 μl of the

DNA, 1X buffer (0.66 M Tris-HCl, pH 7.6, 50 mM $MgCl_2$, 50 mM DTT, 1 mg/ml BSA, 10 mM hexaminecobalt chloride, 2 mM ATP, 5 mM spermidine HCl) 3 μl of 24% PEG and 3 units of T4 polynucleotide kinase. Incubation proceeded at 20°C overnight. *Escherichia coli* TG-1 cells were transformed by electroporation. Plasmids that gave the expected restriction fragments with Asn 1 (Boehringer Mannheim) were sequenced between the Sph 1 and Stu 1 sites. This Sph 1 -Stu 1 fragment was then removed from the mutant pKJ17 plasmid and ligated into a wild-type pKJ17 vector which had not been exposed to the PCR reaction.

2.6 Plasmid DNA preparation

2.6.1 Small scale plasmid preparation

Minipreparations of plasmid DNA were carried out according to the modified alkaline lysis protocol of Birnboim and Doly (1979) and Ish-Horowicz and Burke (1981). *Escherichia coli* strain DH5α bacteria were inoculated into 3 ml of T-broth (Tartof and Hobbs, 1987) containing 100 μg/ml of ampicillin, and grown overnight at 37°C with vigorous shaking. Bacteria from 1 ml of the culture were pelleted by centrifugation at 12,000 g in an Eppendorf tube. The pellet was resuspended in 150 μl of lysis solution (50 mM glucose, 10 mM CDTA, 25 mM Tris-HCl, pH 8.0, 4 mg/ml lysozyme). Following a 5 minute incubation at room temperature, 200 μl of a denaturation solution (0.2 N NaOH, 1% SDS) were added and mixed gently with the lysed bacteria. 150 μl of a high salt solution (3 M potassium acetate, 1.8 M formic acid) were added and mixed

by vortexing. The solution was centrifuged for 5 minutes at 12,000 g at 4°C. The supernatant was extracted once with phenol and chloroform/isoamyl alcohol (24:1 V/V). Plasmid DNA was precipitated by adding two volumes of ethanol to the aqueous phase. The tube was centrifuged 5 minutes at room temperature and the pellet containing the plasmid DNA was resuspended in 50 µl TE containing DNase-free pancreatic RNase. After 15 min at 65°C the DNA was ready for use.

2.6.2 Large scale plasmid preparation

Milligram quantities of highly purified plasmid DNA was prepared using the modified protocol of Marko et al (1982). Bacteria were grown overnight in 250 ml of T-broth containing 100 µg/ml of ampicillin. The culture was centrifuged at 6,000 rpm for 10 minutes at 4°C. The pellet was washed with demineralized water and centrifuged again at 6,000 rpm for 5 minutes at 4°C. The pellet was resuspended in 20 ml of lysis solution (see 2.6.1), and left on ice for 30 minutes. Denaturation solution (40 ml, see 2.6.1) was added and mixed with the bacterial lysate by gently swirling. After storing on ice for 15 minutes, 30 ml of high salt solution (see 2.6.1) were added and mixed. Following a 30 minute incubation on ice, the solution was centrifuged at 10,000 rpm for 20 minutes at 4°C. (All subsequent centrifugations were carried out at 10,000 rpm at 4°C). DNA was precipitated with 2 volumes of ethanol and centrifuged for 15 minutes. The pellet was resuspended in 8 ml of acetate-MOPS (0.1 M sodium acetate, 0.05 M MOPS, pH 8.0), and was precipitated again in ethanol. The centrifuged DNA was dissolved in 3 ml of water and mixed with an equal volume of lithium chloride solution (5 M lithium chloride,

0.05 M MOPS pH 8.0). After a 15 minute incubation on ice and a 20 minute centrifugation, the supernatant was incubated at 65°C for 15 minutes. The solution was centrifuged for another 15 minutes, and the DNA was precipitated from the supernatant, as described above. The pellet was dissolved in 2 ml of acetate-MOPS solution, and the DNA was collected by ethanol precipitation and centrifugation. The pellet was next dissolved in 2 ml of loading buffer (6 M NaCl, 50 mM Tris-HCl, pH 8.0, 10 mM CDTA). Glass powder suspension (20 ml of loading buffer containing 2 g of ground Whatman GF/A filters) was added to the DNA solution, which was gently mixed for 15 minutes. The suspension was filtered under vacuum using a Nalgene 0.2 micron filter. Following a filtration with 100 ml of loading buffer, 24 ml of elution buffer (0.2 M NaCl, 60 mM Tris-HCl, pH 8.0, 10 mM CDTA) were added. The DNA was slowly eluted, and precipitated by adding 3 ml of sodium acetate and two volumes of ethanol. Following a 30 minute centrifugation, the plasmid DNA was resuspended in 1 ml of TE and stored at -20°C.

2.7 RNA Preparation and Analysis

Total RNA was prepared from both P19 cells and HeLa cells according to the method of Auffray and Rougeon (1980).

Northern Blot Analysis. Aliquots of 15 μg of total RNA were ethanol precipitated and resuspended in 20 μl of sample buffer (50% formamide, 10% formaldehyde and 20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, pH 7.0). After heating at 65°C for

10 min and cooling on ice for 5 min, the RNA samples were loaded onto a gel with 1/10 volume of loading buffer (0.1% xylene cyanol, 30% glycerol, 0.05 mg/ml ethidium bromide) and electrophoresed at 60 volts for six hours through 0.9% agarose gels containing 1X RNA gel buffer (20 mM 3-[N-Morpholino]propane-sulfonic acid, 1 mM EDTA, 5 mM sodium acetate, pH 7.0) and 10% formaldehyde. The electrophoretically separated RNAs were transferred by capillary blotting for 16 hours to Hybond-N membranes (Amersham, Oakville, Canada), then the membranes were irradiated with ultraviolet light at 120 mJ using a GS Gene linker UV chamber (BioRad, Mississauga, Canada). Prehybridizations were carried out for 2 hours at 42°C in 50% formamide, 20 X Denhardt's solution, 5X SSPE, 1% SDS and 250 μg/ml denatured herring sperm DNA. Hybridizations were carried out using a boiled, random-primed ^{32}P-labelled DNA insert at 5 x 10^6 cpm/ml at 42°C for 16-24 hours. Filters were washed in several changes of 2X SSC and 0.1% SDS at room temperature for 30 min, followed by several changes of 0.2X SSC and 0.1% SDS at 65°C for 45 min. Finally, hybridizations were visualized by autoradiography, exposing the filters to Kodak XAR-5 film.

DNA used to probe the Northern blot was obtained from Dr.R.Tjian, University of California at Berkeley. The plasmid supplied had been generated by inserting an Nco 1 - Eco R1 fragment from the coding sequence of CTF-1 into the Sma 1 site of pBSK+ (pBluescript, Stratagene Inc.). The Nco 1 site was restored with Klenow, and the fragment was inserted such that it gave an in-frame fusion with 34 amino acids of the lacZ gene. For use as a Northern probe, this plasmid was digested with Bam H1, generating a 1477 bp fragment. The fragment was gel purified, and labelled with the

random primer labelling kit supplied by Pharmacia. The probe was purified from unincorporated nucleotides by precipitation prior to hybridization.

2.8 Protein-DNA Binding Interactions

2.8.1 Band Shift Assay

In the band shift assays (Fried and Crothers, 1981; Garner and Revzin, 1981; Carthew et al 1985; Chodosh et al 1986), 20 μg of nuclear extract were incubated with 3 μg of poly(dI-dC) for 5 minutes at room temperature. To this mixture was added 0.25 ng - 1.0 ng of ^{32}P-labelled, gel purified DNA, and any unlabelled competitor. The reaction was incubated a further 15-30 minutes at room temperature. Binding reactions had a total volume of 20 μl in 10 mM Tris-HCl, pH 8.0, 0.05 M KCl, 10% glycerol and 0.1 M EDTA. Reactions were electrophoresed at 150 volts on 4% (30:1) polyacrylamide gels made up with 0.25% TBE. Gels were dried and subject to autoradicgraphy.

2.8.2 Footprinting Assay

DNase 1 footprinting assays were performed essentially as described by Jones et al (1985). 40 μg of fractionated P19 nuclear extract were incubated with 0.2 ng of ^{32}P-labelled DNA in buffer (12.5 mM Tris-HCl, pH 7.9, 6.25 mM $MgCl_2$, 0.5 mM DTT, 5 μM $ZnSO_4$, 10% (v/v) glycerol, 0.05% (v/v) NP-40, 0.05 M KCl, 2% polyvinyl alcohol) in the presence of 5 μg of poly(dI-dC) for 15 minutes at room temperature. One

minute before digestion, an equal volume of a mixture containing 5 mM $CaCl_2$ and 10 mM $MgCl_2$ was added to the DNA. DNase 1 (Worthington Biochemical Corporation, Freehold, New Jersey) was used at a 1/1000 dilution of a 10 mg/ml stock for control reactions and a 1/50 dilution for test reactions containing extract. The digestion reactions were terminated by the addition of a mixture containing 200 mM NaCl, 20 mM EDTA, 1% SDS, and 250 μg/ml tRNA, extracted once with phenol and chloroform, and precipitated with 2.5 volumes of 95% ethanol. Pellets were rinsed with 70% ethanol and dried before being loaded onto an 8% polyacrylamide gel. Gels were subsequently dried and autoradiographed.

Probes for the footprinting assays were prepared in the following manner. The 303 bp probe of the upper DNA strand was generated by digestion of the plasmid pCAB17prX/T520 with Eco R1 and calf intestinal phosphatase, followed by successive phenol and chloroform extractions, and ethanol precipitation. The DNA fragments were 5'-end labelled with T4 polynucleotide kinase and γ-^{32}P-ATP, then digested with Stu 1. Footprinting of the lower DNA strand was accomplished by the reverse digestion procedure. Labelled probe was purified on an 8% polyacrylamide gel, isolated by elution at 37°C overnight in buffer (0.5 M NH_4OAc, 10 mM $Mg(OAc)_2$, 1 mM EDTA, 0.1% SDS) and subjected to two rounds of extraction with phenol and chloroform. The DNA fragments were precipitated with ethanol and stored in 10 mM Tris-HCl, pH 7.9, 1 mM EDTA at -20°C.

Control sequencing reactions were performed using the double-stranded sequencing protocol of Maxam and Gilbert (1980).

CHAPTER 3

The UAS in Undifferentiated Cells

3.1 Introduction

The ultimate objective of this work was to demonstrate the involvement of the UAS in the developmental transcriptional regulation of the murine *Pgk-1* gene. The initial experiments were concerned with showing that nuclear factors from undifferentiated cells bound to the UAS, delineating their sites of interaction, and demonstrating their involvement in transcription. The results of these experiments are presented in this chapter. The following chapter covers the results of the same sorts of experiments performed using differentiated cell extract, specifically retinoic acid-induced P19 derived cells of neuroectodermal lineage.

Chapters 3 and 4 are divided into a number of sections, each covering a particular line of investigation. The results from each of these sections were used to form the model, presented in Chapter 5, which describes the interaction of factors with the *Pgk-1* promoter.

In Chapter 3, the areas of investigation reported on concern the following. (1) The involvement of UAS binding proteins in X chromosome inactivation. This investigation took the form of a comparison of the binding of nuclear factors, from both male and female cell lines, to the UAS. (2) The determination of the sites of protein-DNA interaction within the UAS. These analyses were carried out using the band shift and DNase 1 footprinting assays. (3) A characterization of the UAS binding activities. One experiment involved an examination of band shift complex stability, by heat denaturation, in order to see if conclusions could be drawn regarding the number of factors which might be involved in the interaction. Another experiment was aimed at determining the specificity of binding to different regions within the UAS, by competition band shift analysis. A third experiment concerned an investigation into the tissue-specific interaction of binding factors with a particular region within the UAS. (4) An examination of the importance of one of the UAS protein binding regions to gene expression. This was assessed by mutating the site of interest and comparing the activities of a linked gene being driven by the *Pgk-1* promoter carrying either the wild-type or mutant binding site, in both transient and stable transfection experiments.

In the band shift assays, there are minor variations in the presence and intensity of retarded complexes. This can be attributed to slight variability in extract preparation, and varying gel exposure times. Consistently reproduced retarded complexes are indicated with arrows.

3.2 Results

3.2.1 Binding to the UAS in male and female cells

All genes on the inactive X chromosome of the mouse are apparently transcriptionally inert, whereas some genes on the inactive X chromosome of humans are transcriptionally active. In humans, the X-linked *PGK-1* allele on the inactive X chromosome is inert.

Riggs' group has determined that protein does not associate with the promoter of the human *PGK-1* gene on the inactive X chromosome (Pfeiffer et al 1990). Therefore, if transcription factors interact with discrete cis-acting DNA elements to influence the course of XCI, it is reasonable to assume that they do so by functioning as activators. By binding to DNA elements associated with particular genes which remain transcriptionally active following XCI, they would prevent those genes from turning off. How these genes are able to retain these transcription factors, while other genes destined to be inactivated are not able to do so, may be key to the mechanism of XCI. Understanding how transcription factors interact with the promoters of X-linked genes to promote, enhance or repress transcription, whether the genes remain active or are inactivated during XCI, is a step closer to understanding the mechanism of XCI.

Towards this end, the binding activities from nuclear extracts of male and female murine EC cell lines were compared, by band shift analysis. The P19, P10(XO), and C86S1A1 cell lines each contain one active X chromosome. P19 is a male cell line with an XY sex chromosome complement, the P10(XO) line is female but only has one X chromosome, and the C86S1A1 line is female, and contains one active and one inactive

X chromosome.

20 μg of nuclear extract from each of these cell lines were incubated with the 303 bp Eco-Stu fragment of the UAS. As seen in Figure 13, two distinct retarded complexes were observed from the extract of each cell line. The doublet which is clearly evident as the faster migrating complex in lanes 3 and 4, also appears to a greater or lesser degree when using P19 nuclear extract (Figure 14). The gel conditions, e.g. pH, exact percentage of acrylamide, may be important for the resolution of this doublet.

Each cell line contained factors capable of binding the *Pgk-1* UAS, and presumably all function the same way to regulate transcription. Therefore, despite the fact that P19 is a male cell line, and all subsequent investigations were carried out using P19 cell extracts, the findings of this thesis would presumably apply to the regulation of the *Pgk-1* allele on the active X chromosome in female cells, and may provide insights into regulation of X-linked genes.

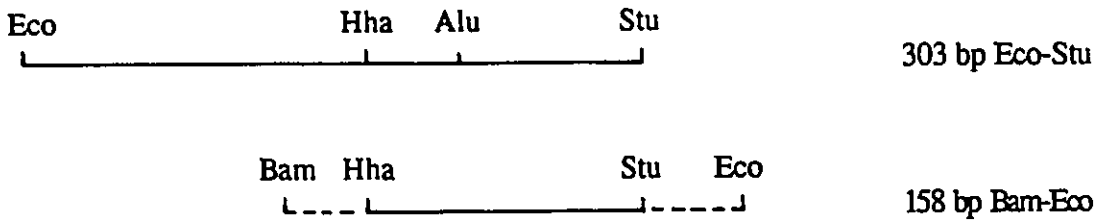
3.2.2 Multiple factors bind to the UAS

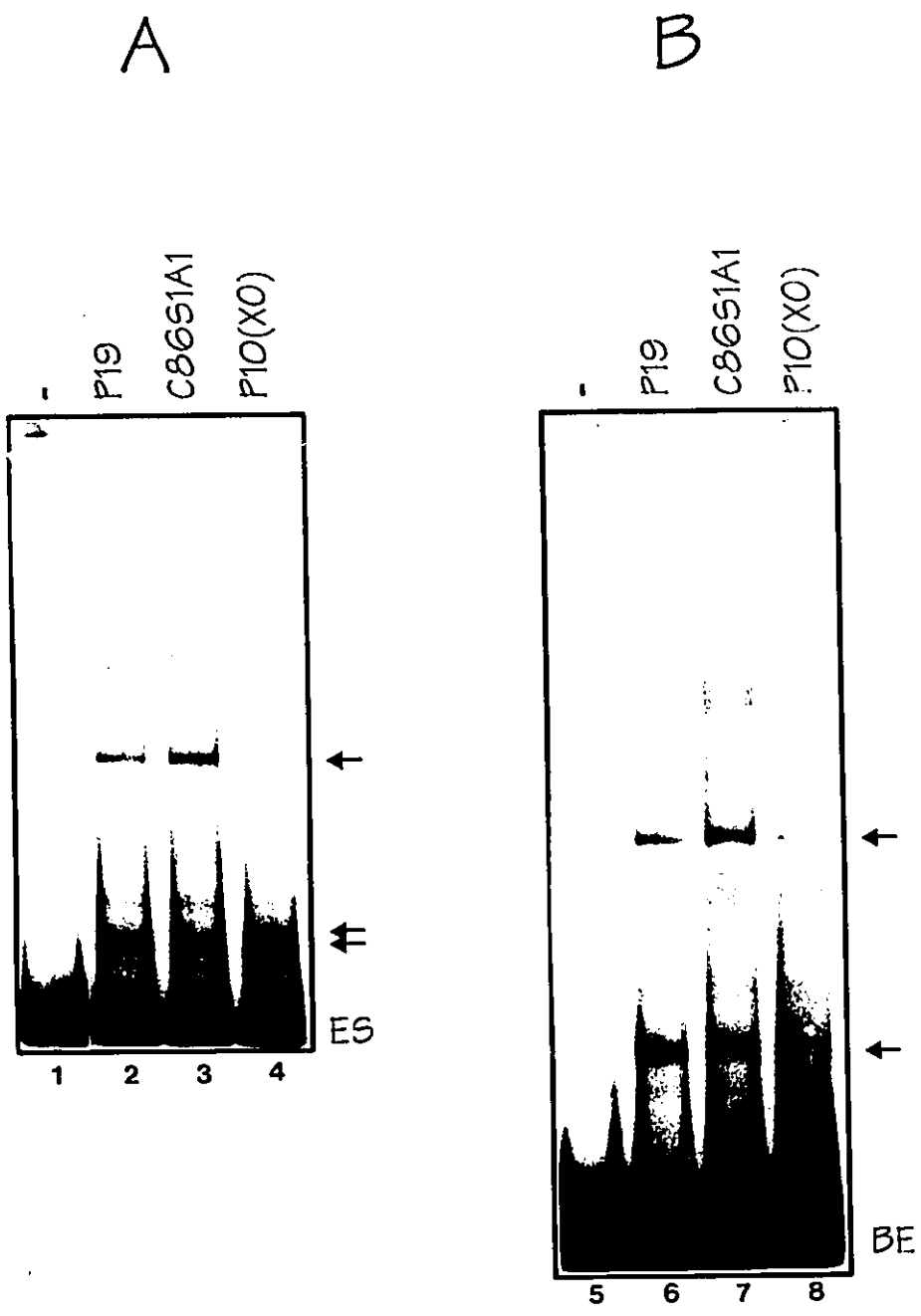
To gain an understanding of how the interaction of protein with the UAS functions to regulate gene expression, the next experiment was aimed at determining whether the factors associating with the UAS form multisubunit associations.

The stability of the retarded DNA-protein complexes associated with the 303 bp Eco-Stu UAS fragment in band shift analyses was examined during a heat denaturation procedure. Different dissociation profiles for the retarded complexes would indicate that the stability of each complex was dissimilar and would suggest that the complexes are

FIGURE 13: Comparison of UAS binding activity in P19, C86 and P10 cells

Binding of nuclear factors to the UAS of the pgk-1 gene was compared by band shift assay between the male P19 cell line and the two female cell lines, C86S1A1 and P10(XO). 20 μg of nuclear extract were incubated with either (A) the 303 bp Eco-Stu, or (B) the 158 bp Bam-Eco UAS fragments. Lanes: 1 and 5, no extract; 2 and 6, P19 nuclear extract; 3 and 7, C86S1A1 nuclear extract; 4 and 8, P10(XO) nuclear extract. Arrows indicate retarded complexes. DNA templates are defined below. Assays were performed a minimum of 10 times.




A
B
-
P19
C8651A1
P10(XO)
ES
BE
1
2
3
4
5
6
7
8

comprised of either varying numbers of the same factor or different proteins altogether.

In the first experiment, labelled Eco-Stu DNA template was incubated with nuclear extract. After 20 minutes at room temperature, the reaction was heated at the indicated temperature for 5 minutes, then loaded onto the gel. In the second experiment, the nuclear extract was first heated for 5 minutes at the indicated temperature, then mixed with the labelled template and incubated 20 minutes at room temperature before loading onto the gel.

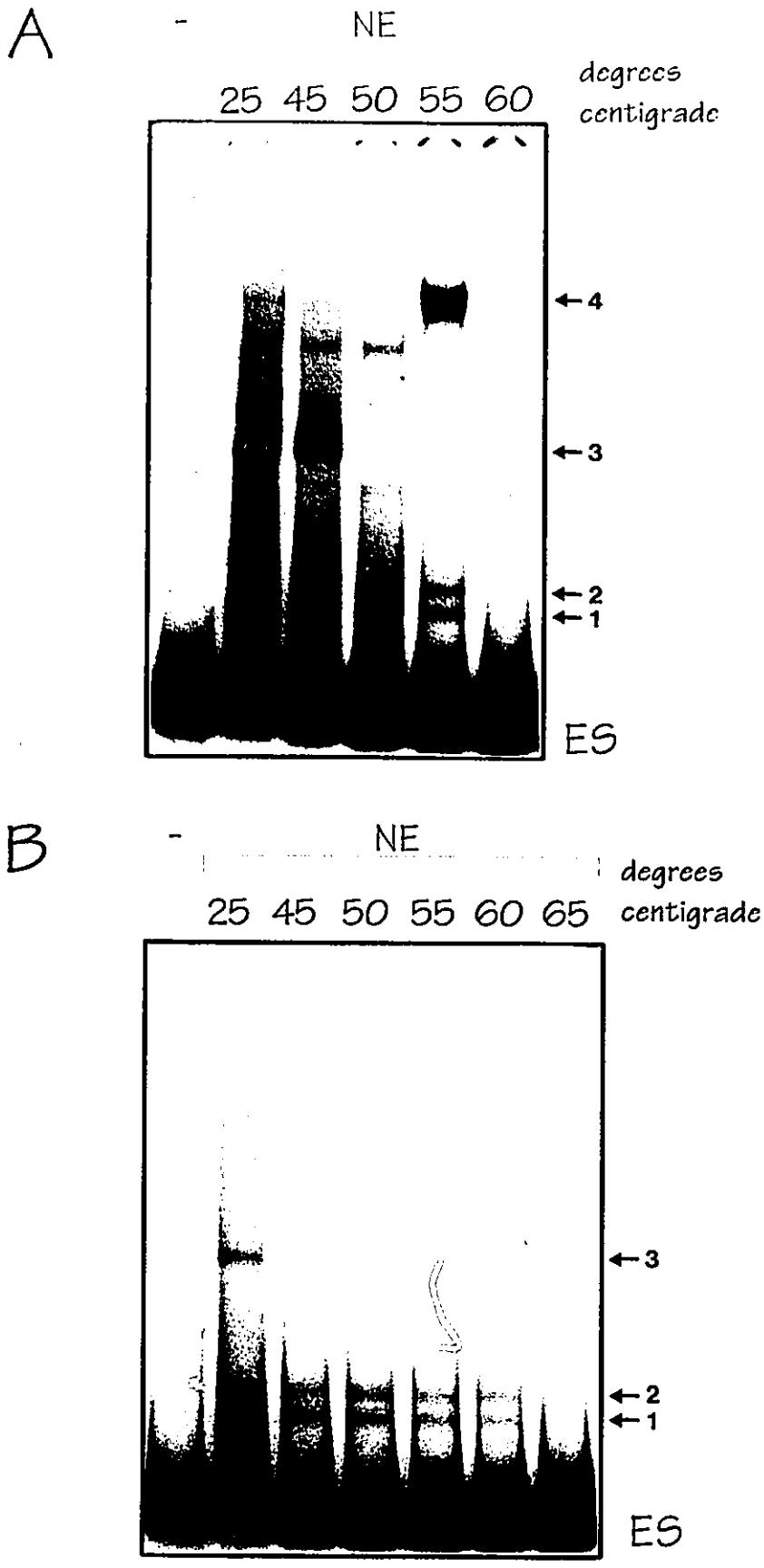
As seen in Figure 14A, when extract was incubated with DNA prior to heating, Complex 3 dissociated from the DNA between 45°C and 55°C, while Complexes 1 and 2 remained associated with the DNA at 55°C but dissociated by 60°C. At 55°C a new, more slowly migrating complex (Complex 4) was observed. Staining of the gel with Coomassie blue stain did not reveal an accumulation of protein at that site, which would indicate a clump of non-specifically binding protein. Therefore Complex 3 and the Complexes 1 and 2 have different heat stabilities, and Complex 4 represents a specific association of factors with the UAS which does not occur unless the factors remain bound to the DNA during the heating process.

As seen in Figure 14B, when extract was heated prior to incubation with the DNA, Complex 3 was unable to form if the extract had been incubated above 25°C, while Complexes 1 and 2 were able to form until the extract was heated above 60°C. Therefore, even when the extract was heated prior to binding, Complex 3 and the Complexes 1 and 2 revealed different heat stabilities.

The different heat dissociation profiles for Complex 3 and the Complexes 1 and

FIGURE 14: Stability of UAS transcription complex

The heat stability of the DNA:protein complexes was examined in band shift assays. **(A)** Labelled 303 bp Eco-Stu fragment was incubated with 20 μg of P19 nuclear extract for 20 minutes, then heated at the temperatures indicated for 5 minutes before loading onto the gel. Complexes 1, 2 and 3 are those normally seen in band shift assays using the Eco-Stu restriction fragment as template. Complex 4, lane 5, was only observed when the reaction was heated to 55 °C, following binding. **(B)** Labelled Eco-Stu fragment was incubated with 20 μg of P19 nuclear extract which was first heated at the temperatures indicated for 5 minutes. Lane 1 of (A) and (B) contains no extract. Assays were reproduced four times.


A
-
NE
25 45 50 55 60
degrees
centigrade
4
3
2
1
ES
B
-
NE
25 45 50 55 60 65
degrees
centigrade
3
2
1
ES

2 in both experiments suggests that more than a single polypeptide is involved in these binding interactions. The complexes are either composed of different numbers of the same protein, or different types of proteins, and therefore are not affected in the same way by the heat treatment.

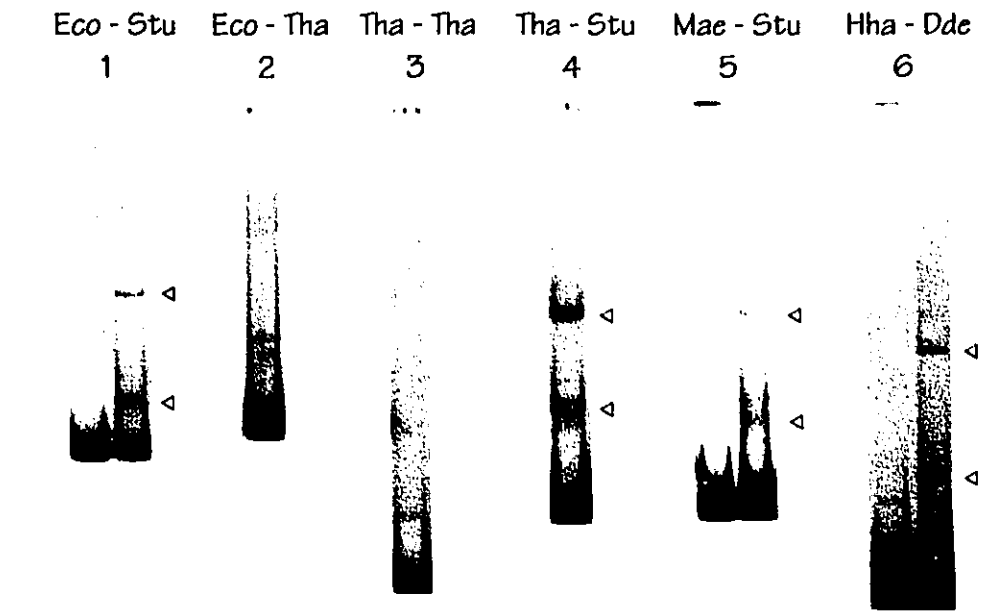
Also, comparing Figure 14A and 14B, it is apparent that the protein in Complex 3 is more resistant to heat denaturation when bound to DNA (14A) than when not bound (14B). When pre-bound to the DNA before heating, Complex 3 does not dissociate until after 45°C has been reached; however, when not pre-bound to the DNA prior to heating, the DNA-protein complex is unable to form if the extract has been heated above 25°C. This result suggests that the binding activity represented by Complex 3 involves multiple factors that may interact and thereby stabilize the binding.

3.2.3 Sites of interaction

Having determined that nuclear factors interact with the UAS of the *Pgk-1* gene, the region was subdivided into smaller fragments, to more clearly delineate the binding sites, by band shift assay. The Tha 1 restriction fragment was used, as Tha 1 digests the Eco-Stu fragment into three DNA fragments of reasonable sizes for band shift analysis (Figure 15). Only the Tha 1 restriction fragment which is most proximal to the site of transcription initiation supported binding. Additional digestion analyses of the proximal region narrowed down the binding site to the 54 bp Mae 3 - Dde 1 fragment. The known consensus binding sites in the region within, and in the immediate vicinity of, the Mae3 - Dde 1 stretch are outlined in Figure 16.

FIGURE 15: Sites of protein:DNA interaction

Band shift assays on six different restriction fragments. Arrows indicate protein:DNA complexes. The first lane of the Eco-Stu, Mae-Stu and Hha-Dde band shifts contained no nuclear extract. All other lanes contained 20 μg of P19 nuclear extract. Restriction fragments 2, 3 and 4 cover the whole of the UAS, which was digested with Tha 1. The origins of the restriction fragments, and a summary of their binding activity is outlined. Binding to the UAS, as observed by band shift analysis, was restricted to the proximal Mae 3-Dde region. Assays were performed a minimum of two times.


Eco - Stu
Eco - Tha
Tha - Tha
Tha - Stu
Mae - Stu
Hha - Dde
1
2
3
4
5
6

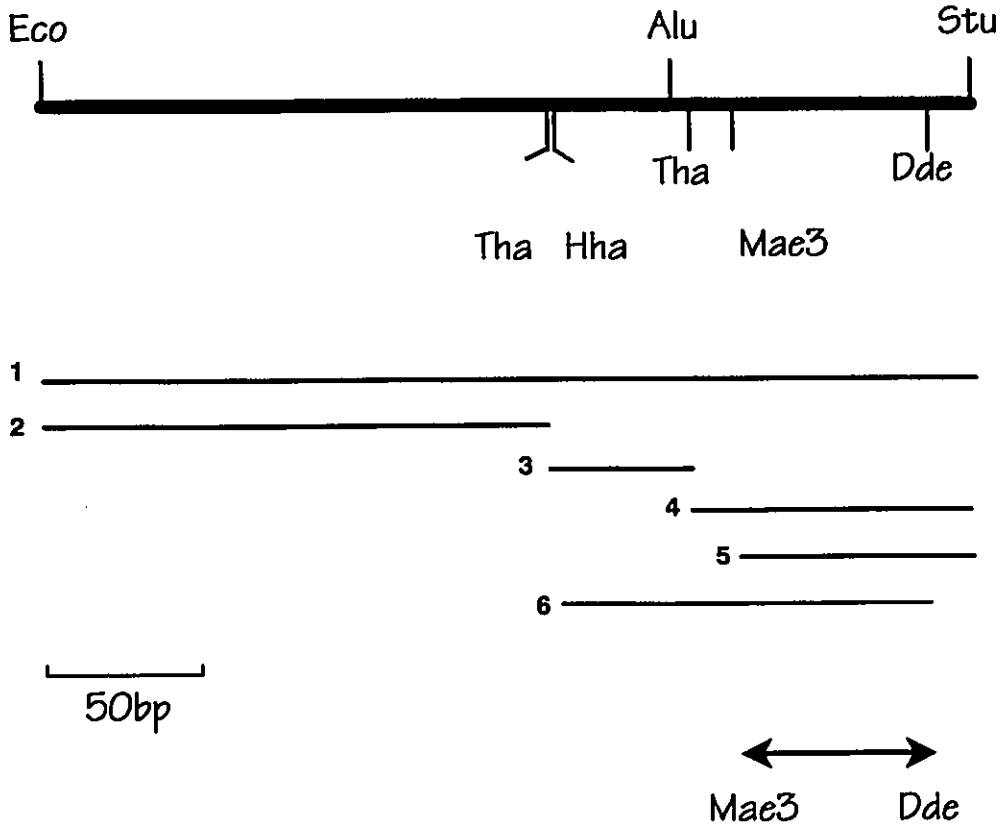
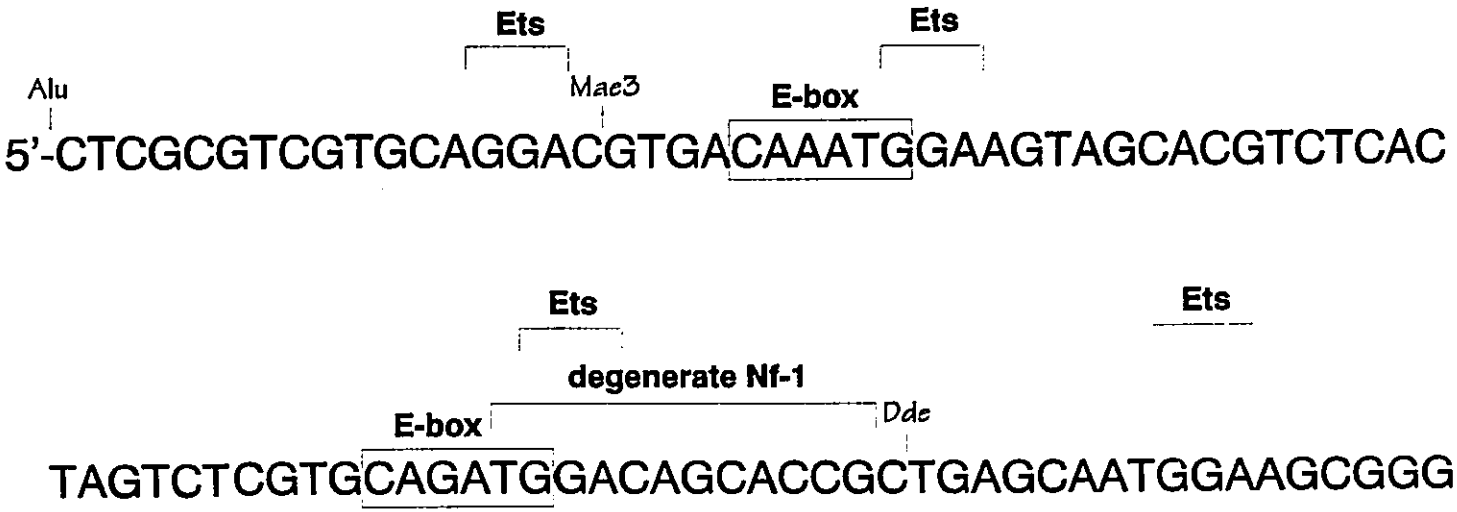

Eco
Alu
Stu
Tha
Dde
Tha
Hha
Mae3
Binding Activity
1
2
3
4
5
6
+
-
-
+
+
+
50bp
Mae3
Dde

FIGURE 16: Consensus protein binding sites within the proximal region of the UAS

The sequence of the UAS region between the Alu 1 and Stu 1 restriction sites is presented. Protein binding was found to occur within the Mae 3 and Dde 1 restriction sites (Figure 15). The E-box, Ets, and degenerate Nf-1 consensus binding sites are delineated.



E-box: 5'-CANNTG-3'

Ets: 5'-GGA-3'

Nf-1: 5'-TGGC/A(N)5GCCA-3'

degenerate Nf-1: 5'-TGGA(N)4ACCG-3'

R2 footprints

Clearly defined binding sites help in the identification of putative transcription factors, simplify mutation studies for expression analyses, and provide more accurate probes for screening. Footprinting studies were therefore performed in order to more accurately delineate the site(s) of protein-DNA interaction.

Initially, footprinting experiments were performed using total nuclear extract from P19 cells, but a footprint to the region was never detected. When fractionation of the nuclear extract over heparin-agarose was carried out, there were never sufficient quantities left after band shift analyses to test adequately in footprint analyses. Sufficient quantities of fractionated material were supplied by Dr.St-Arnaud of McGill University. The nuclear extract was from P19 cells which had been exposed to retinoic acid for four days then fractionated over wheat germ agglutinin and sephacryl-300 columns. Footprinting studies were performed using this material.

Each fraction was first examined for binding activity to the UAS by band shift analysis, then the fractions which contained binding activity were used in the footprinting assay. By incubating the 303 bp Eco-Stu fragment of the UAS with 10 μl of each fraction in a band shift assay, it was determined that fractions 14-19 from the NE-17 preparation would be used in the footprinting studies (data not shown).

As seen in Figure 17, a footprint was detected to a region which was subsequently termed R2. Delineated in Figure 18 - with reference to the known consensus binding sites - 11 bases on the upper DNA strand, and 26 bases on the lower strand were protected from DNase 1 digestion. The footprint encompassed an E-box (5'-CANNTG-3'), the

FIGURE 17: R2 produces a footprint

DNase 1 footprint of the R2 region of the UAS. **(A)** The upper DNA strand of the 303 bp Eco-Stu fragment, labelled at the Eco site. Lanes: 1, control reaction, no extract; 2 and 3, 18 and 36 μg of Fraction 18, NE-17. **(B)** The lower DNA strand, labelled at the Stu site. Lanes: 1-4, Maxam and Gilbert sequencing reactions; 5, control reaction, no extract; 6, 36 μg of Fraction 18, NE-17.

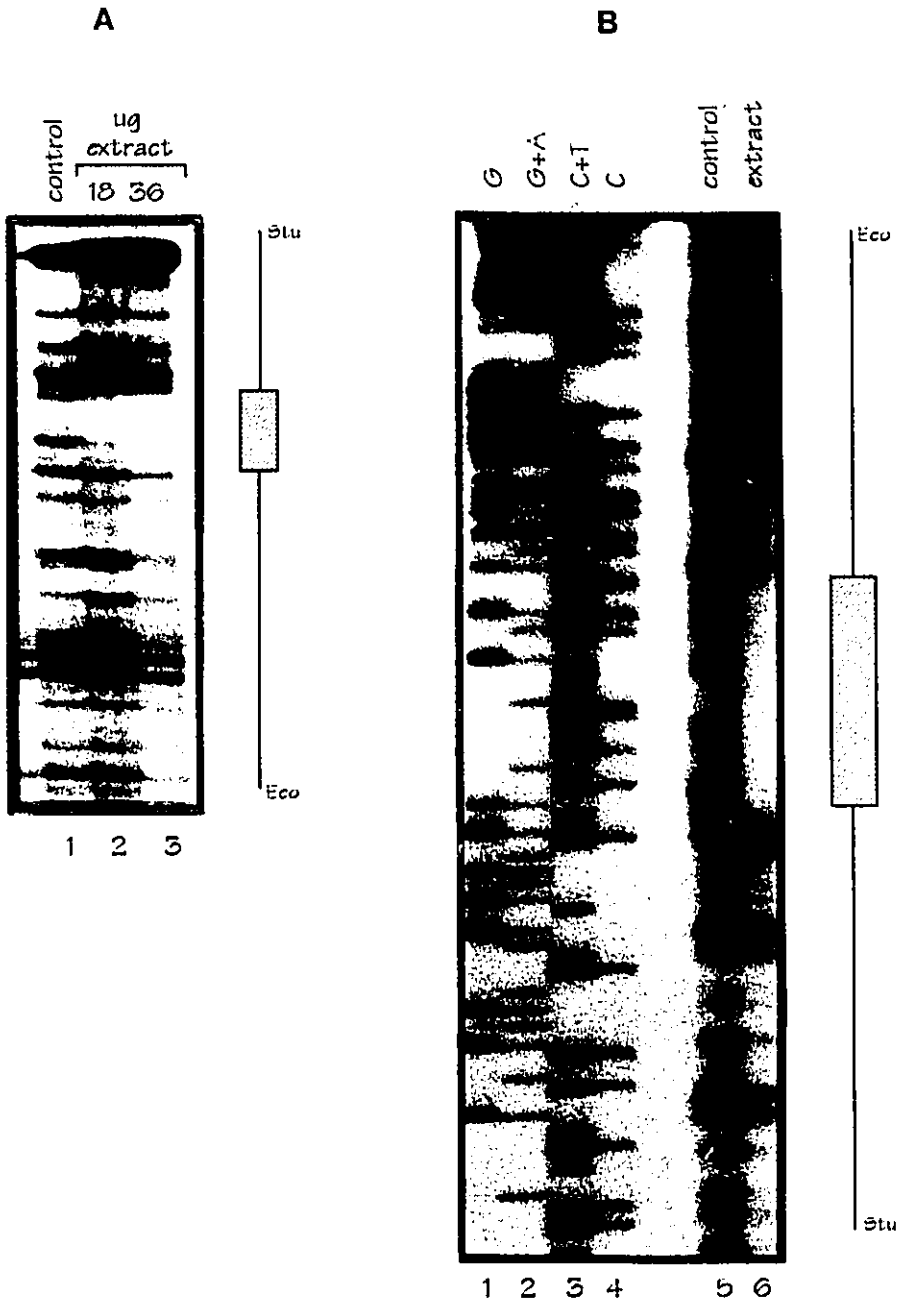
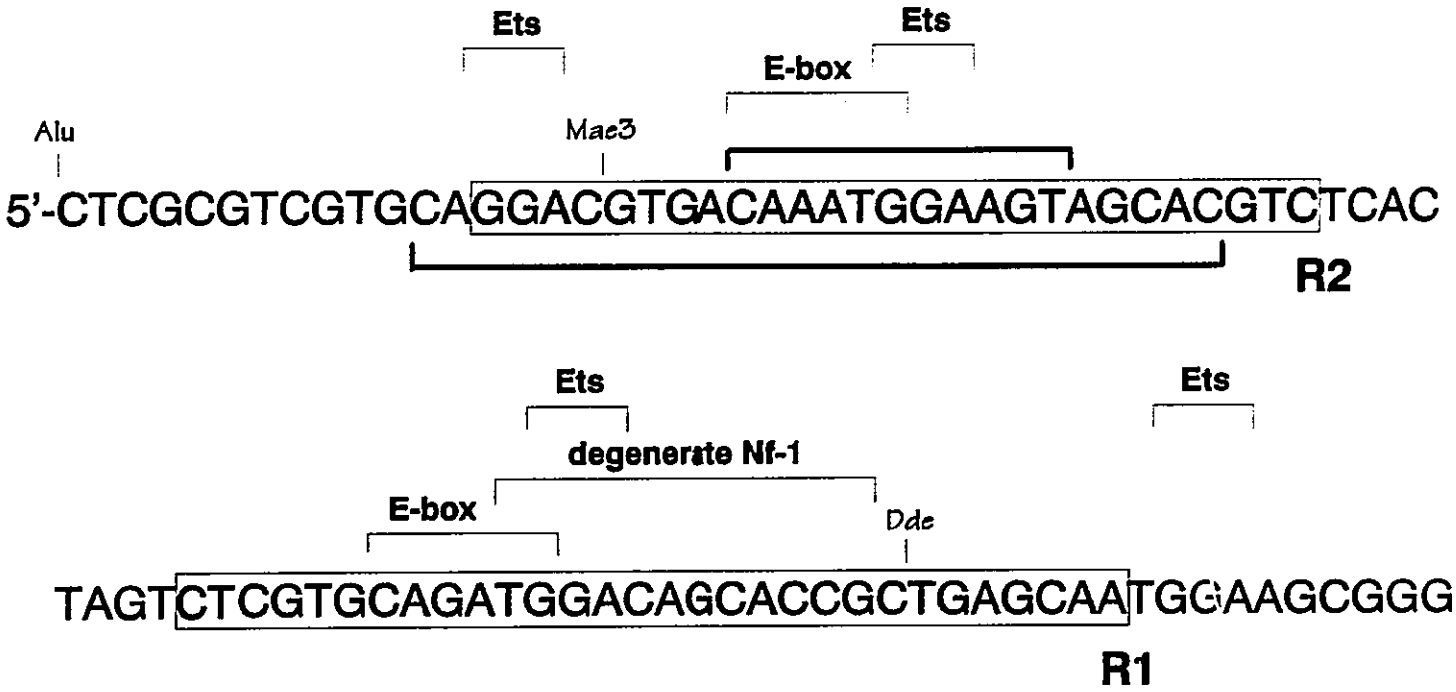

A
control
ug extract
18 36
Stu
Eco
1 2 3
B
G
G+A
C+T
C
control
extract
Eco
Stu
1 2 3 4 5 6

FIGURE 18: R2 footprinted sequence

Delineation of the R2 DNA sequence protected from DNase 1 digestion. In reference to the known protein consensus binding sites in the region, the bases which were protected from DNase 1 digestion are shown. The dark bracket above the sequence designated R2 delineates the 11 bases on the upper DNA strand which were protected. The dark bracket under the sequence delineates the 26 bases on the lower strand which were protected. The sequences included in the R1 and R2 oligonucleotides are enclosed within the shaded boxes.



E-box: 5'-CANNTG-3'

Ets: 5'-GGA-3'

Nf-1: 5'-TGGC/A(N)5GCCA-3'

degenerate Nf-1: 5'-TGGA(N)4ACCG-3'

consensus binding site for helix-loop-helix proteins, and two Ets protein binding sites (5'-GAA-3'). The most 5' of the two Ets binding sites, however, did not fall within the Mae3 - Dde 1 region which had previously been determined to be the only region within the UAS to support protein binding (Figure 15).

The footprint was obtained using an extract from retinoic acid treated P19 cells. As retinoic acid induces the differentiation of P19 cells, the binding activity detected at R2 is from a differentiated cell extract. However, as will be seen in the next chapter where a study was conducted which compares the binding of factors to the UAS in untreated and retinoic acid treated P19 cell extracts, no detectable binding difference was observed at R2. Therefore the footprint obtained using the retinoic acid treated extract is likely the same as a footprint which would be obtained using untreated P19 nuclear extract prepared in the same manner.

One explanation for the fact that R2 would not footprint with unfractionated P19 nuclear extract is that the R2 binding activity has a higher affinity for a protein which *in vivo* would bind other DNA specifically.

Binding at R1

Figure 18 shows that a cluster of consensus binding sites occurs in a region of the Mae3 - Dde 1 sequence 3' to the R2 binding site. Consequently an oligonucleotide was generated to this region, with the sequence 5'-CTCCGTGCAGATGGACAGCACCGCTGAGCAA-3', extending past the Dde 1 restriction site in order to centre the consensus binding sites within the fragment. At the

same time an oligonucleotide was generated to the R2 region (5'-GGACGTGACAAATGGAAGTAGCACGTC-3'), again centering the consensus binding sites which were 3' to the Mae3 site.

Each oligonucleotide was used in a band shift assay. The DNA was incubated with 20 μg of P19 nuclear extract, and 20 μg of nuclear extract with a 10-fold molar excess of unlabelled self DNA. The results are shown in Figure 19. Both R1 and R2 were shown to contain specific protein-binding activity.

Therefore, within the 54 bp Mae3 - Dde 1 region of the UAS are at least two sites of protein interaction, designated R1 and R2, R2 delineated by band shift and footprint analyses and R1 by band shift analysis.

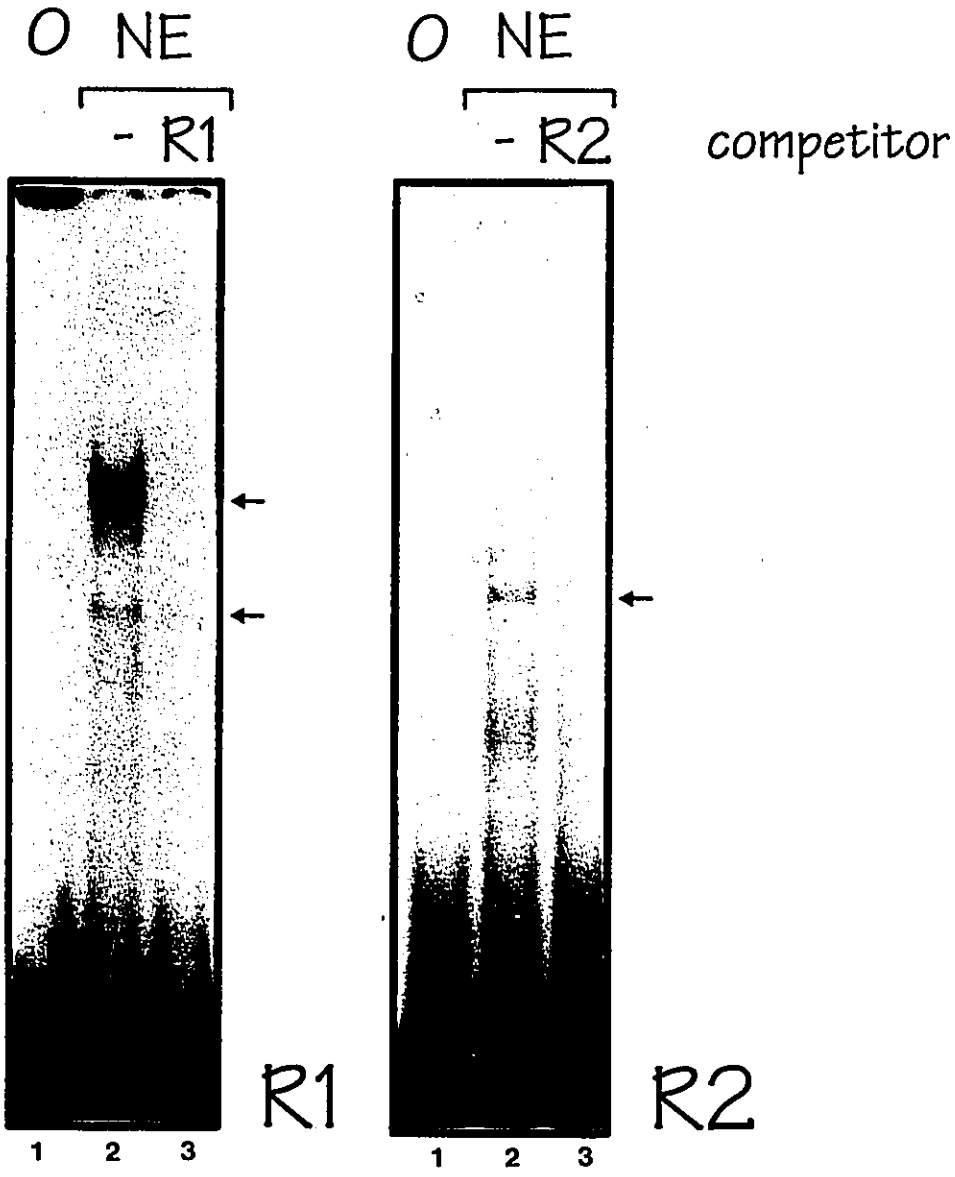
3.2.4 Binding activity at R1 and R2 differs

Both R1 and R2 contain an E-box binding site (R1: 5'-CAGATG-3'; R2: 5'-CAAATG-3') and Ets binding sites (5'-GGA-3'). In order to determine whether the same factor or subset of factors interacts with both R1 and R2, a competition band shift experiment was performed.

Nuclear extract was incubated with labelled R2 oligonucleotide template and mixed with unlabelled R1 or R2 oligonucleotide prior to addition to the extract mix. When the same DNA was used as competitor and template, the retarded bands were expected to disappear. If the same factor(s) were able to interact with both R1 and R2, with the same binding affinity, then when the unlabelled R1 DNA was used as competitor, the retarded bands would also be expected to disappear.

FIGURE 19: R1 and R2 binding activities

R1 and R2 were used in band shift assays and found to bind nuclear factors. In each experiment, lane 1 contains no protein, lanes 2 and 3 contain 20 μg of P19 nuclear extract, and lane 3 contains unlabelled self DNA, at 10-fold molar excess relative to the labelled DNA, used as competitor to show specificity of binding. Assay was performed twice.


O NE
- R1
O NE
- R2
competitor
R1
R2
1 2 3
1 2 3

When labelled R2 template was incubated with unlabelled R1 DNA, the retarded bands remained (Figure 20), indicating that R1 was unable to compete for R2 binding factor(s). It was therefore concluded that R2 and R1 either (1) interact with different factors, or (2) interact with the same binding activity but the affinity of interaction is greater for R2 than for R1.

3.2.5 R2 mutation affects binding and expression

To determine if the E-box and Ets binding sites within the Mae3 - Dde 1 sequence of R2 were involved in the protein interaction, the sites were mutated and the effect on binding activity was examined. This site was also mutated to determine if the R2 site was important for *Pgk-1* gene expression in P19 cells, and hence of possible importance in the developmental transcriptional regulation of PGK-1.

The mutation was generated using oligonucleotide primers and PCR technology to introduce an 8 bp replacement of sequence in the centre of the footprint. The sequence 5'-CAAATGGA-3' was changed to 5'-AATTAATT-3', thereby destroying the E-box and Ets binding sites (Figure 21).

The 303 bp Eco-Stu fragment, incorporating either the wild-type or the mutant R2 sequence, was incubated with P19 nuclear extract in a band shift assay. As shown in Figure 22, the mobility of the retarded complexes was changed by the mutation, indicating that binding was affected.

FIGURE 20: Binding activity at R1 and R2 differs

Labelled R2 oligonucleotide was mixed with unlabelled R1 or R2 oligonucleotide at a 20-fold molar excess, and incubated with 20 μg of P19 nuclear extract. Lanes: 1, free R2 oligonucleotide probe; 2-4, 20 μg P19 nuclear extract; 2, no competitor; 3, 1 μg unlabelled R2 DNA; 4, 1 μg unlabelled R1 DNA. Assay was performed once.

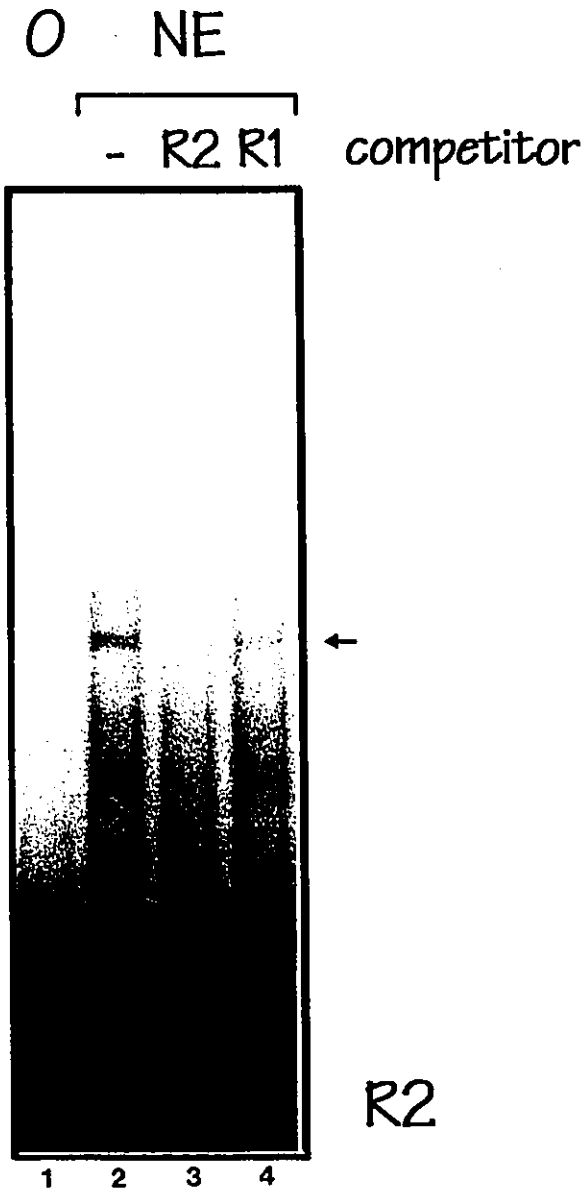

0
NE
- R2 R1
competitor
R2
1 2 3 4

FIGURE 21: Sequence of the R2 mutation

The sequences of the R2 region and the mutated R2 region are delineated. Lines above and below the R2 sequence represent bases protected in DNase 1 footprinting. 8 bp in the wild-type sequence were replaced in the mutant; replacement bases are bolded in the mR2 sequence.

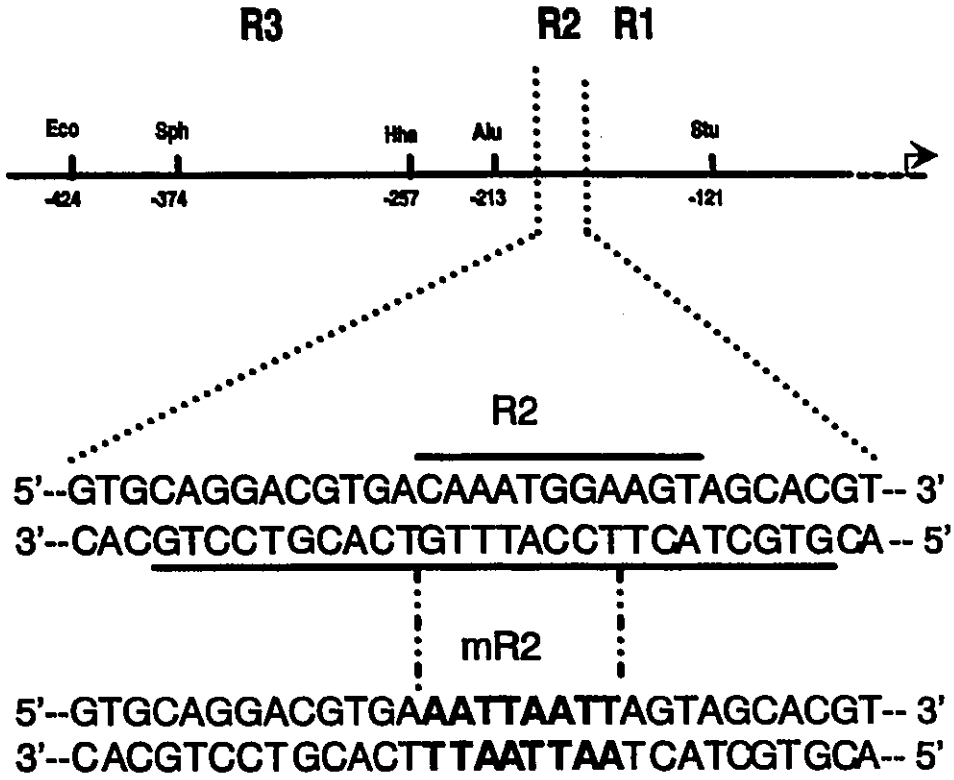
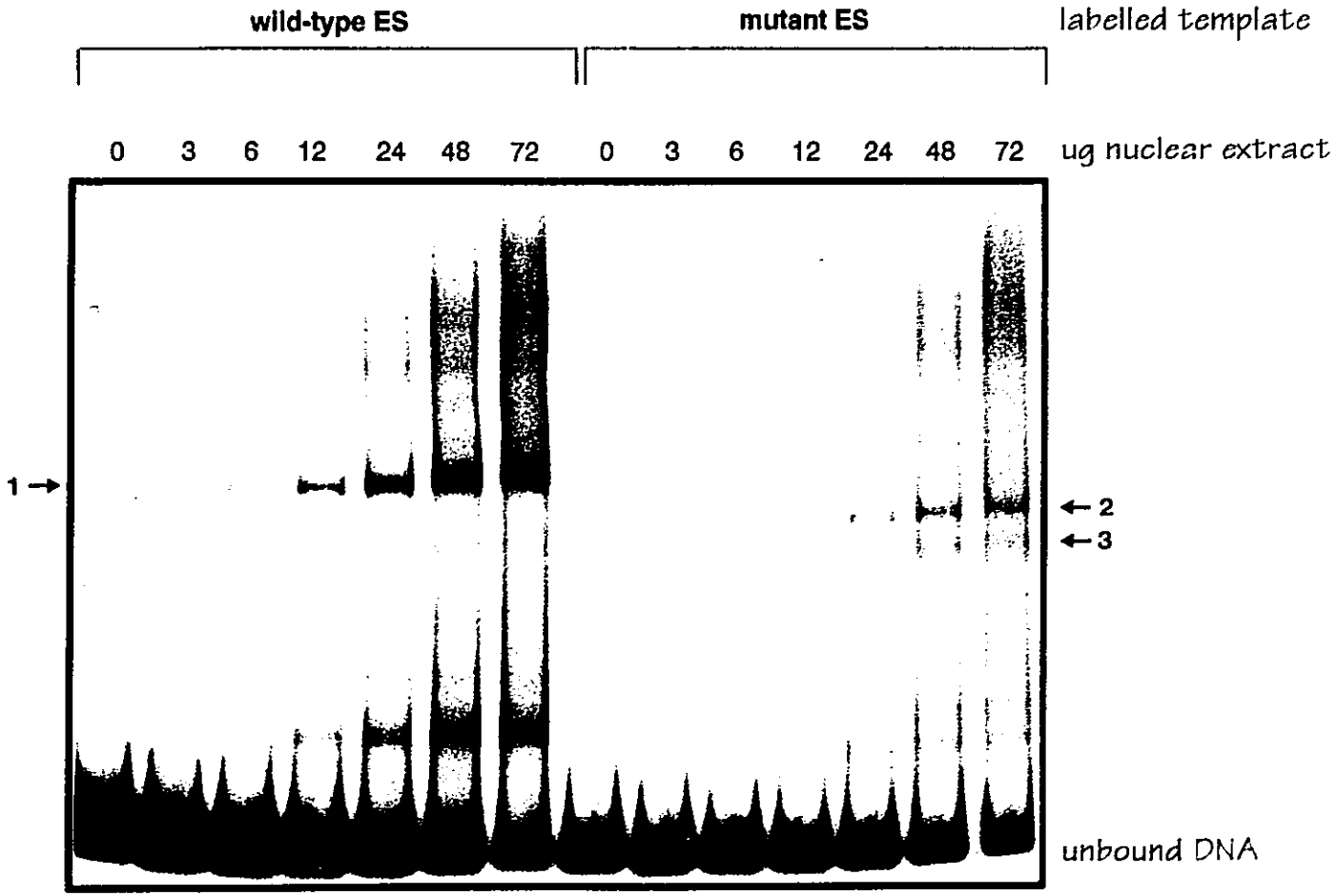

R3
R2
R1
Eco
Sph
Hha
Alu
Stu
-424
-374
-257
-213
-121
R2
5'--GTGCAGGACGTGACAAATGGAAGTAGCACGT-- 3'
3'--CACGTCCTGCACTGTTTACCTTCATCGTGCA-- 5'
mR2
5'--GTGCAGGACGTGAAATTAATTAGTAGCACGT-- 3'
3'--CACGTCCTGCACTTTAATTAATCATCGTGCA-- 5'

FIGURE 22: R2 mutation leads to altered binding

In a band shift assay, either the wild-type Eco-Stu fragment or the Eco-Stu fragment containing the mutant R2 site, was incubated with increasing amounts of P19 nuclear extract. Complex 1 was observed when nuclear extract was incubated with the wild-type Eco-Stu sequence: Complexes 2 and 3 were observed when nuclear extract was incubated with the Eco-Stu fragment carrying the mutation in the R2 region. Assay was reproduced a minimum of five times.


wild-type ES
mutant ES
labelled template
0 3 6 12 24 48 72 0 3 6 12 24 48 72
ug nuclear extract
1
2
3
unbound DNA

mR2 footprint

Closer examination of the effect of the mutation on the interaction of proteins with the R2 site was carried out by footprint analysis. The same fractionated retinoic acid treated P19 nuclear extract was used here as was used to obtain the original R2 footprint.

The footprint of the mutated R2 region within the 303 bp Eco-Stu fragment (Figure 23), showed a continued protection of 11 bases on the upper DNA strand, but a reduced area of protection on the lower strand, from 26 to 12 bases, compared to the footprint of the unmutated R2 sequence (delineated in Figure 24). Despite the fact that almost the entire sequence on the upper strand which footprinted on the wild-type R2 DNA was changed by the mutation, it was interesting to note that it was still protected on the mutated sequence. One explanation was that specific binding was occurring to the lower DNA strand, which resulted in a non-specific protection of the upper DNA strand from DNase 1 digestion during footprinting. As 12 bases on the lower DNA strand were no longer protected in the mutant sequence, it appears that more than one DNA binding factor is able to interact at the R2 site, which the mutation changes.

The R2 site involvement in gene expression

The mutation in the R2 region was next used to determine the importance of the R2 site on *Pgk-1* gene expression.

The CAT construct containing the *Pgk-1* promoter harbouring the R2 mutation, termed mR2, was both transiently and stably transfected into P19 cells. Cells were transiently transfected with 5 μg of either the plasmid carrying the CAT gene being

FIGURE 23: DNase 1 footprint of mR2 region

Using the DNase 1 footprinting technique, 40 μg of Fraction 18, NE-17 were incubated with the end-labelled 303 bp Eco-Stu fragment containing the mutated R2 sequence. Lanes 1 and 3 are control lanes with no added extract. Lanes 2 and 4 contain fractionated extract.

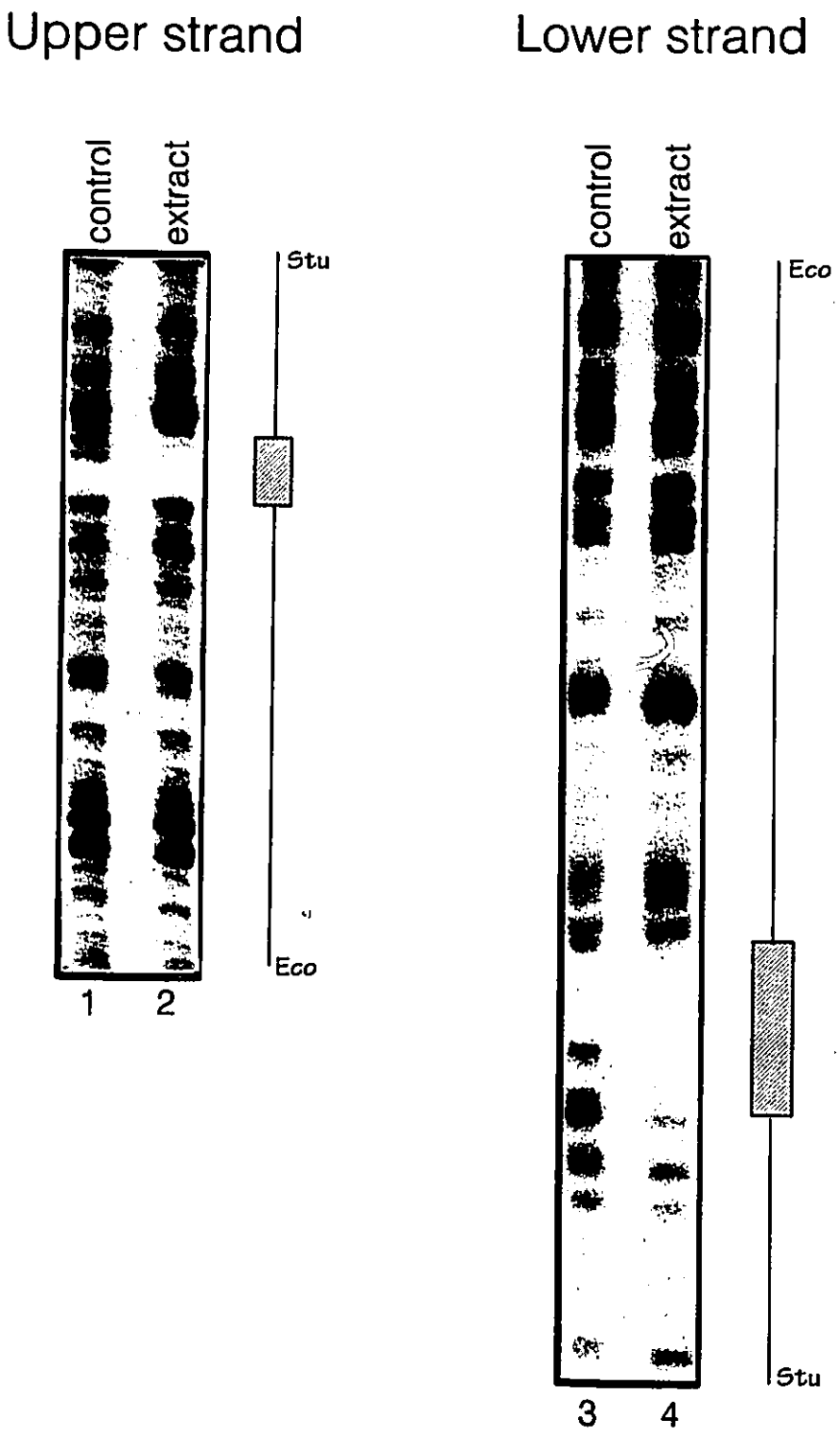
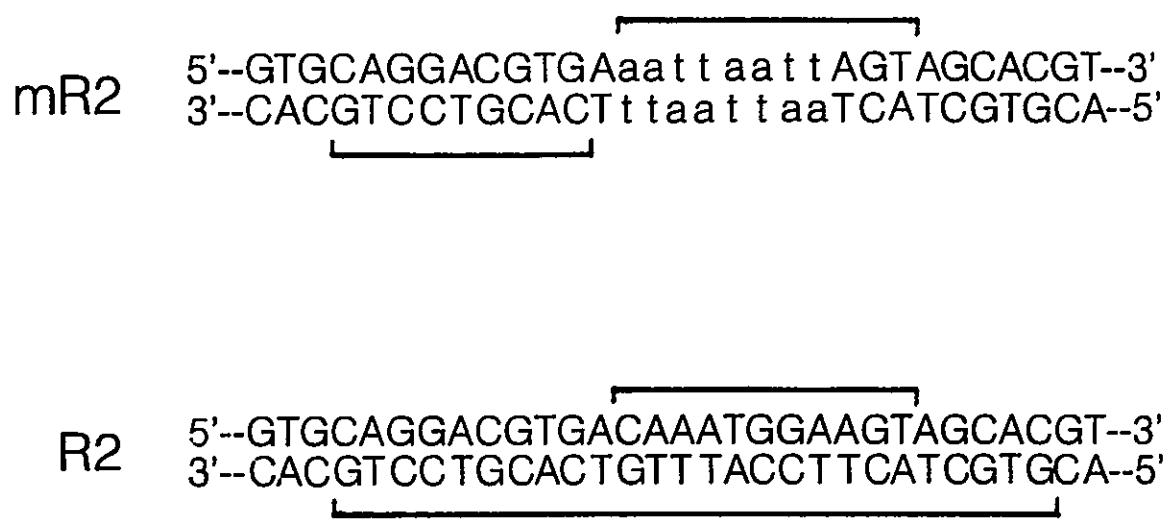

Upper strand
Lower strand
control
extract
Stu
Eco
1
2
control
extract
Eco
Stu
3
4

FIGURE 24: mR2 footprinted sequence

The sequences of the R2 and mR2 regions are shown, with bases protected from DNase 1 digestion being delineated by lines above and below the sequences. The mR2 footprint covers 11 bases on the upper DNA strand and 9 bases on the lower DNA strand, whereas the R2 footprint covers 11 bases on the upper DNA strand and 26 bases on the lower DNA strand.


mR2
5'--GTGCAGGACGTGAaatt aatt AGTAGCACGT--3'
3'--CACGTCCTGCACTttaat taaTCATCGTGCA--5'
R2
5'--GTGCAGGACGTGACAAATGGAAGTAGCACGT--3'
3'--CACGTCCTGCACTGTTTACCTTCATCGTGCA--5'

driven by the wild-type *Pgk-1* promoter (pKJ17), or by the plasmid mR2, and 5 μg of pKJ7-lacZ(*Pgk*-lacZ) for a control of transfection efficiency. Cells were stably transfected with 5 μg of either test plasmid, 5 μg of the control plasmid, 2 μg of B17 and 2 μg of *Pgk*- puromycin.

In transiently transfected P19 cells there was no effect of the mutation on CAT expression (see Figure 25A). In stably transfected P19 cells, however, mutation of the R2 site resulted in a 5-fold reduction in CAT expression levels (see Figure 25B).

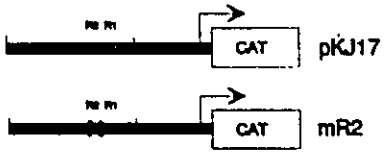
The major differences between stably and transiently transfected cells are that (1) the copy number of the transfected DNA is much higher in transiently transfected cells where the DNA remains episomal, and (2) the stably transfected DNA has nucleosomes associated with it (Archer et al 1992). Assuming that the R2 binding activity functions to open up the chromatin by bending and twisting the DNA and perhaps displacing nucleosomes, then a mutation in the binding site leading to an alteration in binding activity would have more noticeable effects in stable transfections where the chromatin is prevented from assuming the more open configuration, than in transient transfections where the DNA is relatively efficiently transcribed and assumes an open chromatin configuration (Archer et al 1992; Laemmli et al 1992).

In conclusion, the R2 site is important for *Pgk-1* gene expression in P19 cells, and may function by altering chromatin structure.

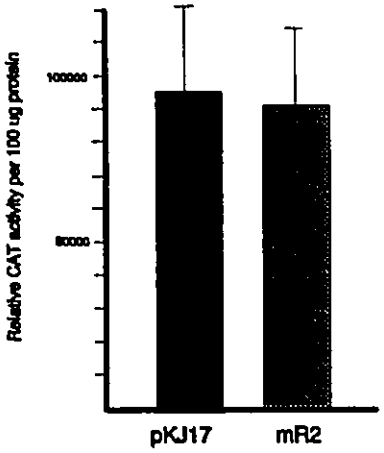
FIGURE 25: Effect of R2 mutation on expression in P19 cells

(A) Transient transfection of either the wild-type pgk-1 UAS in the pKJ17 plasmid, or the pgk-1 UAS carrying the R2 mutation in the mR2 plasmid. 5 μg of test plasmid were co-transfected with 5 μg of the control plasmid pgk-lacZ.

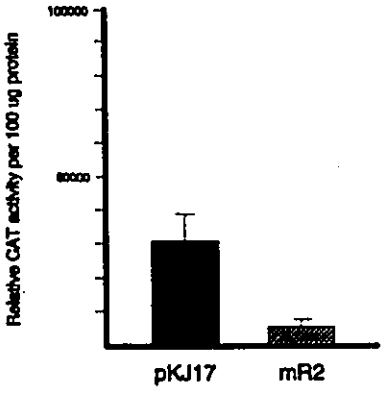
(B) Stable transfection of the same test plasmids. 5 μg of test plasmid were co-transfected with 5 μg of control plasmid, 2 μg of B17 and 2 μg of pgk-puromycin. Both transfections were performed in duplicate. Error bars represent thte standard error of the mean. P=0.05.


CAT
pKJ17
CAT
mR2

A


Relative CAT activity per 100 ug protein
100000
80000
pKJ17
mR2

B


Relative CAT activity per 100 ug protein
100000
80000
pKJ17
mR2

3.2.6 R1 binding is tissue- or species-specific

To summarize the results obtained so far regarding the binding of nuclear proteins from P19 cells to the UAS, it has been determined that at least two regions, R1 and R2, bind factors, that the binding interaction at the two sites differs, and that R2 likely interacts with more than one DNA binding protein. It is unclear, however, which sites within R2 are important for this interaction. The following experiments focus on identifying the sites of protein interaction within R1.

Nf-1 binding

In the human *PGK-1* promoter at about the same position relative to transcription initiation as the murine R1 site, there is a consensus Nf-1 binding site (5'-TGGA/C$(N)_5$GCCA-3'). This site has been *in vivo* footprinted in human cells (see Chapter 1). The murine R1 region contains a degenerate Nf-1 binding site with the sequence 5'-TGGA$(N)_4$ACCG-3'. The fact that the site is degenerate, combined with the fact that undifferentiated cells do not appear to contain Nf-1 family members (McQuillan et al 1991; Speck and Baltimore, 1987; Goyal et al 1990; Nakshatri et al 1990) (see Chapter 1), makes it unlikely that R1 interacts with any Nf-1 related factor in P19 cells. To examine the possibility, however, the next set of experiments sought to determine if P19 cells contain any transcription factors able to interact with the Nf-1 consensus sequence.

In the first experiment, P19 RNA was probed in a Northern blot using an Nf-1 family member cDNA probe (CTF-1) to determine if P19 cells produce any RNA for the

Nf-1 family of transcription factors. In the second experiment, P19 nuclear extract was tested by band shift analysis for binding factors able to interact with a known Nf-1 consensus binding site.

In the Northern blot experiment, total RNA from either P19 or HeLa cells was hybridized with a 1500 nucleotide cDNA probe of the CTF-1 transcription factor (see Chapter 2 for details). As seen in Figure 26A, HeLa cells produced abundant message for CTF-1, which is alternately spliced and produces two transcripts (McQuillan et al 1991), whereas P19 cells produced barely detectable quantities of message which hybridized with the CTF-1 probe.

In the band shift assay, a 26 bp oligonucleotide containing a strong Nf-1 binding sequence (sequence in Table 1, Materials and Methods) was incubated with either P19 or HeLa cell nuclear extract. As seen in Figure 26B, strong binding activity to the Nf-1 template was only observed with the HeLa extract and weak binding activity with the P19 extract.

It appears, therefore, that Nf-1 related transcription factors are present at low concentrations in P19 cells, as in other EC cells. The binding activity associated with R1 is therefore unlikely to be an Nf-1 family member.

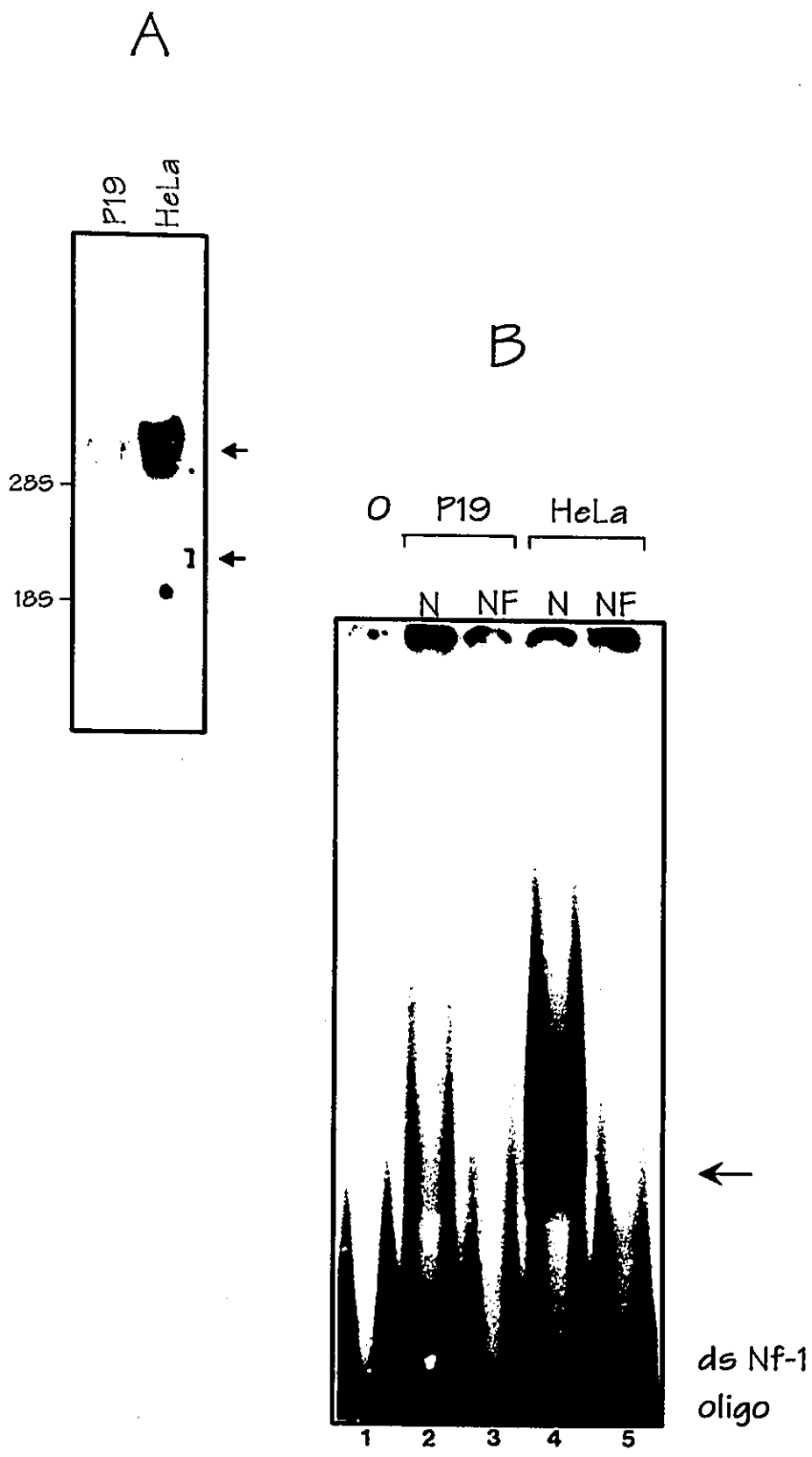
E-box binding

In addition to the degenerate Nf-1 consensus binding sequence, the R1 site contains an E-box consensus sequence and an Ets binding site. Although the E-box sequence is not conserved in the human *PGK-1* promoter, an examination of its potential

FIGURE 26: Nf-1 transcription factors in P19 cells

(A) Northern blot. 15 μg of total RNA from P19 and HeLa cells were probed with a 1500 bp CTF-1 cDNA. The 28S and 18S RNA were used as markers. Arrows indicate the presence of two signals in the HeLa RNA lane, and one signal in the P19 RNA lane. To confirm that equal concentrations of RNA from each cell line were probed, the RNA in the gel was stained with ethidium bromide and visualized on a transilluminator.

(B) Band shift assay. P19 and HeLa nuclear and nuclear-free cytoplasmic cell extracts were incubated with the labelled Nf-1 oligonucleotide. Lane 1: no protein; lanes 2 and 4: 20 μg of P19 and HeLa nuclear extract (N), respectively; lanes 3 and 5: 50 μg of P19 and HeLa nuclear-free (NF) cytoplasmic extract, respectively. The arrow marks the position of the faint P19 signal and the strong HeLa signal. Assay was performed once.


A
P19
HeLa
28S
18S
B
0
P19
HeLa
N
NF
N
NF
1
2
3
4
5
ds Nf-1
oligo

involvement in protein interaction at the R1 site in the murine UAS was carried out.

A mutation of the E-box was incorporated into an oligonucleotide (mR1), with the E-box sequence 5'-CAGATG-3' changed to 5'-GTGATG-3' (see Table 1, Materials and Methods or Figure 19 for the complete oligonucleotide sequence). P19 and HeLa nuclear extract were tested with the R1 and mutant R1 oligonucleotides in a band shift assay.

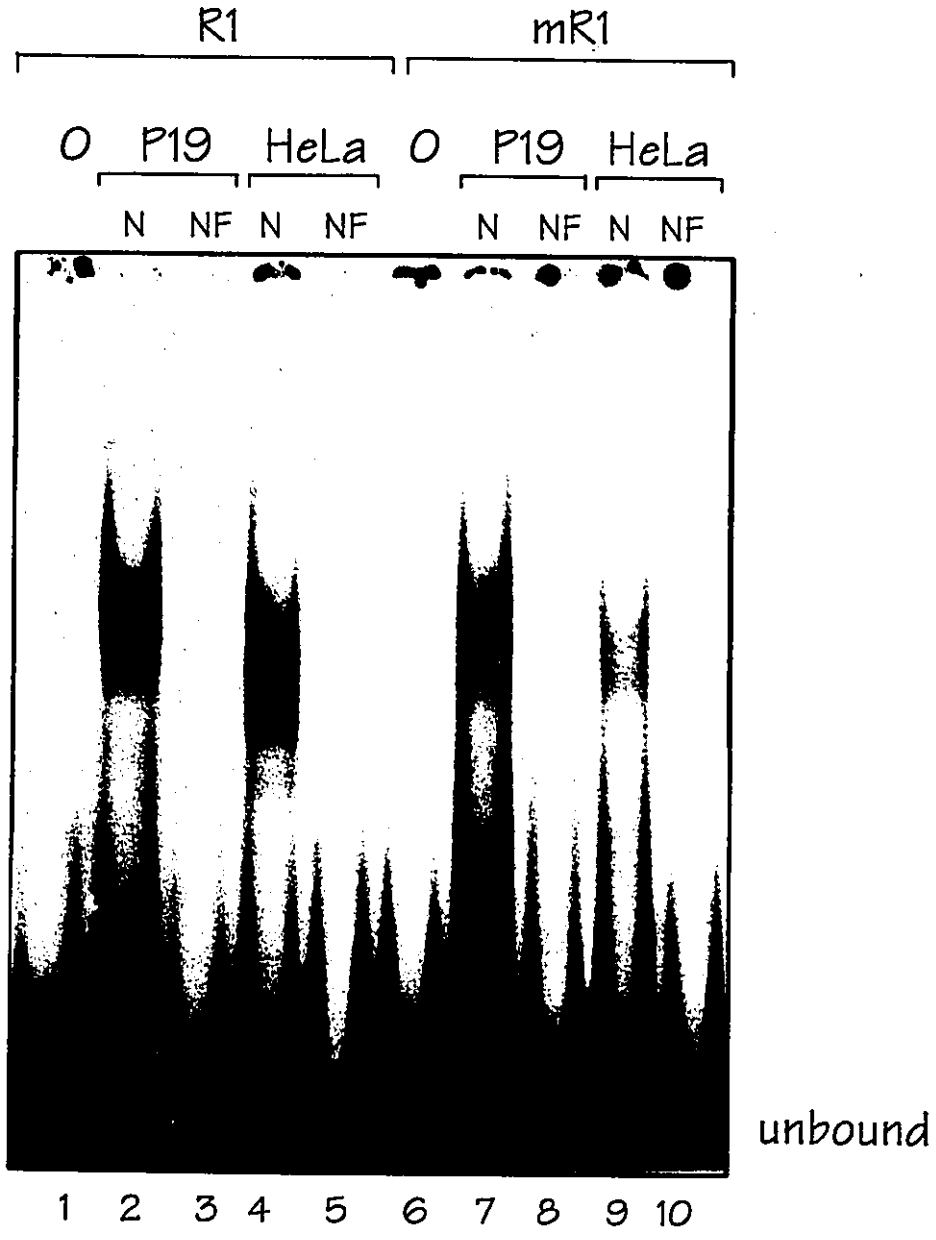
As seen in Figure 27, the mobility of the retarded complex formed using P19 nuclear extract differed from the mobility of the retarded complex formed using HeLa nuclear extract. The HeLa complex was less retarded than the P19 complex. Mutation of the E-box sequence in the R1 DNA did not appear to effect binding of P19 factor(s), but distinctly effected binding of HeLa factor(s). The HeLa complex formed on the mR1 DNA gave a much weaker signal than on R1, and had a different mobility and banding pattern.

Therefore the E-box consensus sequence in the R1 region does not appear to bind factor(s) from P19 cells, but does from HeLa cells. This result suggests that the interaction of nuclear factors with the R1 site is tissue-or species-specific.

As this result had significant implications with regard to the original hypothesis that the UAS has an important role to play in the developmental transcriptional regulation of *Pgk-1*, a more thorough examination of the interaction of P19 and HeLa nuclear factors with R1 was undertaken. This included a more complex evaluation of the binding interaction by band shift analysis, and an examination of the DNA binding activities, observed by band shift analysis, following fractionation of each extract over heparin-agarose. Details of the column fractionation are described in Materials and Methods. The

FIGURE 27: Mutant R1 E-box affects binding in HeLa cells

In a band shift assay the binding activities of the wild-type and mutant R1 oligonucleotides were compared. The R1 sequence was 5'-CTCGTGCAGATGGACAGCACCGCTGAGCAA-3' and the mR1 sequence was 5'-AGCTTCTCGTGGTGATGGACAGCACCGCTGAGCAAA-3'. The bolded, underlined bases are different between the two oligonucleotides. R1 or mR1 were incubated with P19 or HeLa cell nuclear or nuclear-free cytoplasmic extracts. Lanes 1 and 6: no protein; lanes 2 and 7: 20 μg of P19 nuclear extract (N); lanes 4 and 9: 20 μg of HeLa nuclear extract (N); lanes 3 and 8: 50 μg of P19 nuclear-free cytoplasmic extract (NF); lanes 5 and 10: 50 μg of HeLa nuclear-free cytoplasmic extract (NF). Assay was performed once.


R1
mR1
0
P19
HeLa
0
P19
HeLa
N
NF
N
NF
N
NF
N
NF
unbound
1
2
3
4
5
6
7
8
9
10

results presented here are from the column #1 fractionation.

The results of the fractionation experiments are presented in Figures 28 and 29. In each reaction, 10 μl of fractionated material were incubated with the R1 oligonucleotide. Of all the P19 nuclear extract fractions tested, only Fraction 7 contained a significant level of DNA binding activity. This fraction contained the highest level of protein of any fraction from the 0.24 M KCl elution. Of the HeLa nuclear extract fractions tested, a number showed binding activity, but of the same retarded complex, apparently due to a rather broad fractionation profile. What is of interest is that for both the P19 and HeLa extracts, one retarded complex was observed, and in both extracts this complex had the same mobility. This suggests that the DNA binding component of R1 in both P19 and HeLa cells is the same, and that any observed differences in binding using total nuclear extracts are due to additional, non-DNA binding proteins which associate with the DNA-binding protein.

The results of the complex evaluation of the binding differences at R1 between P19 and HeLa nuclear extracts are presented in Figure 30. The gel was electrophoresed for a much longer time than the previous ones, resulting in a clearer separation of retarded complexes. Lanes 1-4 show the banding pattern which results when increasing quantities of HeLa nuclear extract are incubated with the R1 oligonucleotide. The arrows indicate five complexes. Lanes 5 and 6 contain two different concentrations of HeLa nuclear-free cytoplasmic extract, which forms one retarded complex. Lanes 7-10 show the banding pattern which results when increasing quantities of P19 nuclear extract are incubated with the R1 oligonucleotide. The circles indicate four complexes. Lanes 11 and

FIGURE 28: P19 cell heparin-agarose column fractions

A band shift assay testing the binding activity of the P19 nuclear extract fractions from the first heparin-agarose column described in Chapter 2. Fractions 3, 7 and 11 had the highest levels of protein from the 0.1 M, 0.24 M and 0.6 M KCl elutions, respectively. 20 μg of P19 nuclear extract (lane 1), and 10 μl of each fraction, except 7 and 8 (2 μl), were used in the reactions. The arrow points to the single retarded complex observed in Fraction 7. Assay was performed twice.

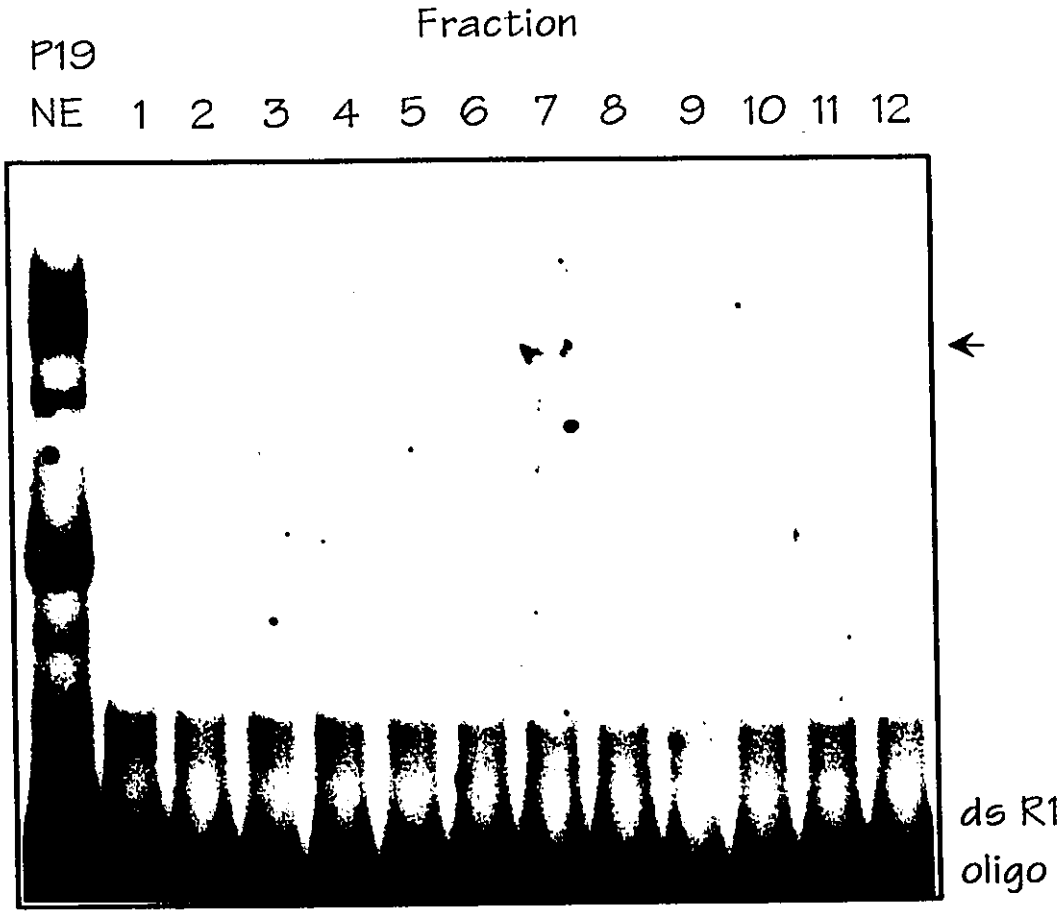

Fraction
P19
NE 1 2 3 4 5 6 7 8 9 10 11 12
ds R1
oligo

FIGURE 29: HeLa cell heparin-agarose column fractions

HeLa cell nuclear extract was fractionated over heparin-agarose and each fraction was tested for binding activity in a band shift assay. 20 μg of nuclear extract or 10 μl of each fraction were incubated with the R1 oligonucleotide. The first lane contained P19 nuclear extract, the second lane contained HeLa nuclear extract, and the remaining lanes, the HeLa fractions. The arrow indicates the position of the one retarded complex present in the fractionated material. Assay was performed once.

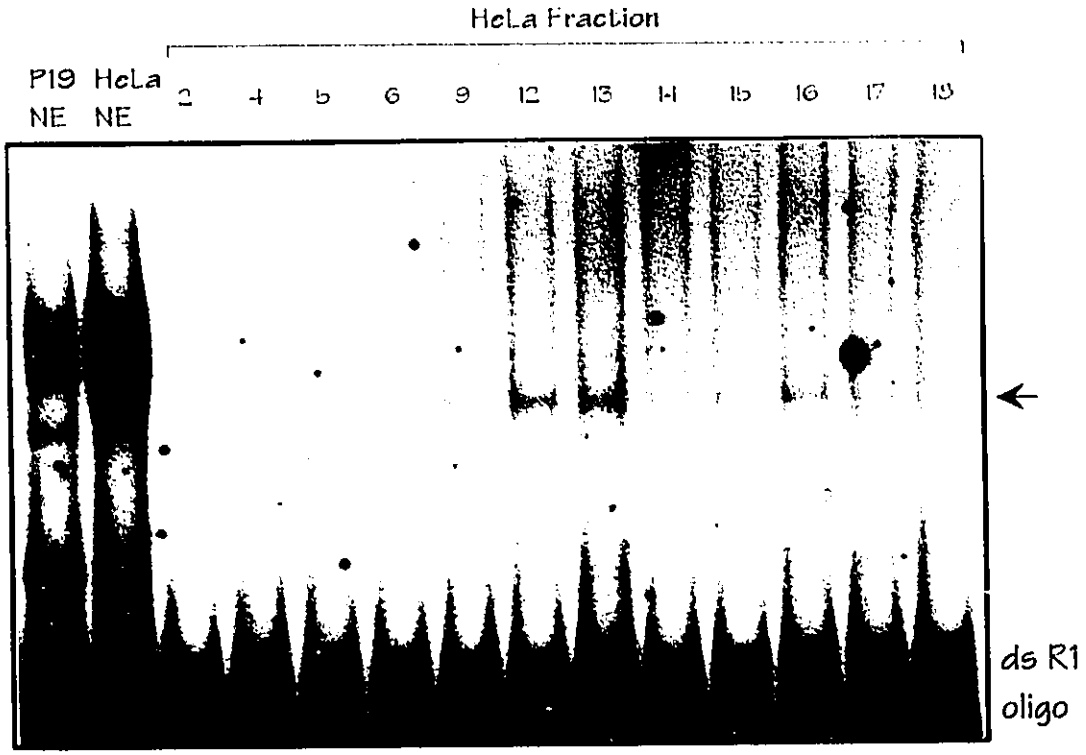
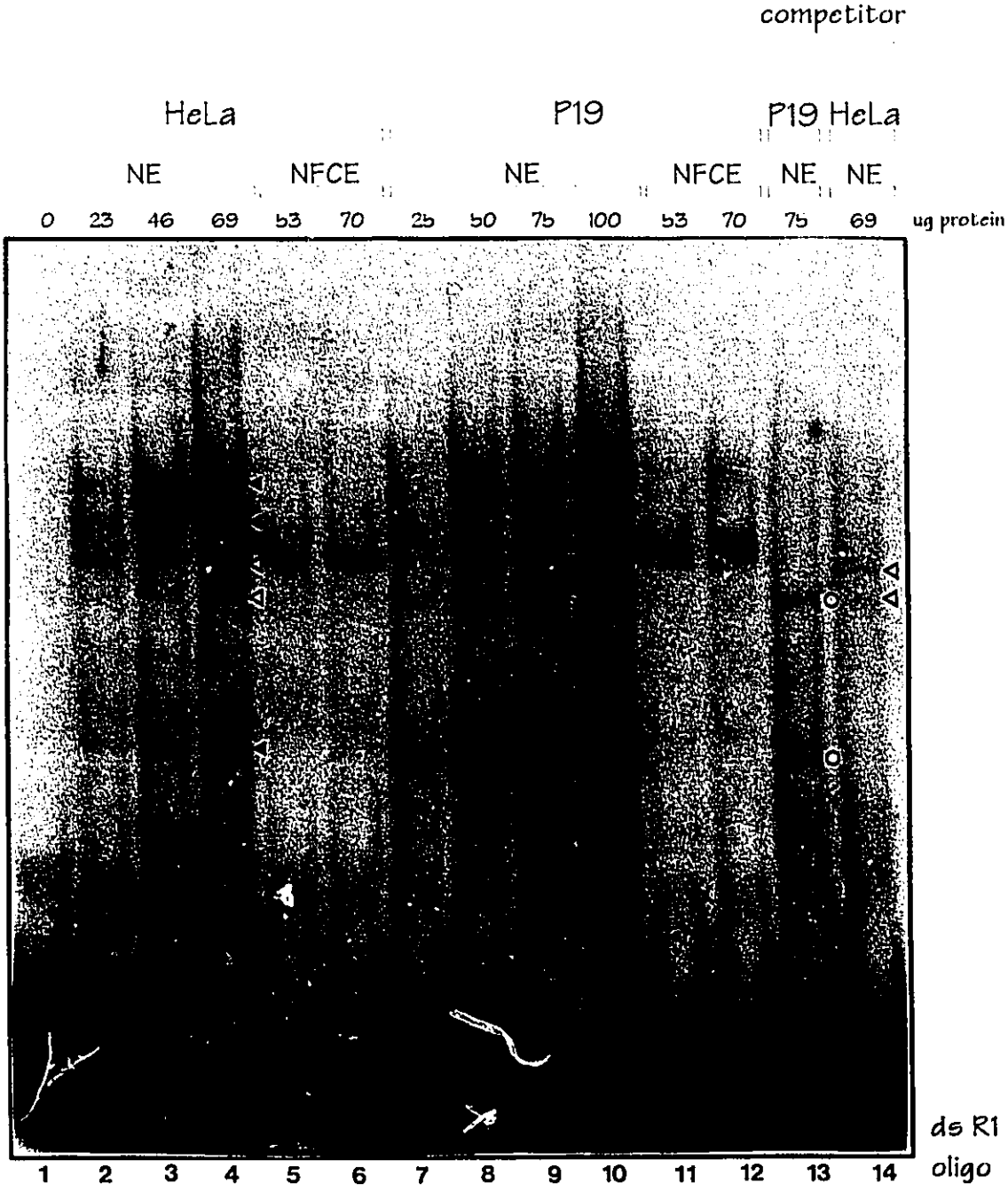

HeLa Fraction
P19 NE
HeLa NE
2
4
5
6
9
12
13
14
15
16
17
19
ds R1 oligo

FIGURE 30: R1 binding activity is tissue- or species-specific

A comparison of R1 binding activities was made by band shift analysis between P19 and HeLa, nuclear and nuclear-free, cell extracts. A titration of all extracts was performed to see if higher order complexes formed in the presence of increasing quantities of protein. Roughly equivalent amounts of P19 and HeLa nuclear extract were also incubated with labelled template in the presence of unlabelled template in order to determine which complexes contained specific binding activities. Lanes: 1-4, HeLa nuclear extract; 5 and 6, HeLa nuclear-free cytoplasmic extract; 7-10, P19 nuclear extract; 11 and 12, P19 nuclear-free cytoplasmic extract; 13, P19 nuclear extract and unlabelled R1 competitor DNA; 14, HeLa nuclear extract and unlabelled R1 competitor DNA. Solid triangles indicate specific protein:DNA complexes in the HeLa nuclear extract, hollow triangles indicate non-specific protein:DNA complexes. Solid circles indicate specific protein:DNA complexes in the P19 nuclear extract, hollow circles indicate non-specific protein:DNA complexes. Assay was performed three times.


competitor
HeLa
P19
P19 HeLa
NE
NFCE
NE
NFCE
NE
NE
0 23 46 69 53 70 25 50 75 100 53 70 75 69
ug protein
ds R1
oligo
1 2 3 4 5 6 7 8 9 10 11 12 13 14

12 contain two different concentrations of P19 nuclear-free cytoplasmic extract, which forms two retarded complexes. Lanes 13 and 14 contain roughly equal quantities of either P19 or HeLa nuclear extract, respectively, and excess unlabelled R1 oligonucleotide as competitor DNA. The retarded complexes which are observed in lanes 13 and 14 therefore represent non-specific protein-DNA binding interactions, and are indicated by open circles and triangles in these lanes and also in lanes 4 and 9. Therefore, HeLa nuclear extract incubated with R1 forms three specific retarded protein complexes in band shift assays, and P19 nuclear extract forms two. Also the nature of the interactions appears to be somewhat different, as the specific P19 complexes are very diffuse whereas the specific HeLa complexes are quite sharp.

Combining these results with those of the fractionation experiments, it appears that the tissue-or species-specific binding observed to the R1 region of the UAS is due to non-DNA binding proteins which are able to form complexes with the DNA-binding protein(s).

3.3 Conclusions and Discussion

The experimental results presented in this chapter support the original hypothesis, discussed in Chapter 1, that the UAS of the murine *Pgk-1* gene interacts with nuclear factors. The two sites of protein-DNA interaction which were identified within the UAS, R1 and R2, were delineated by band shift analyses (R1) and *in vitro* footprint analysis (R2). It remains a possibility that there are more sites of protein interaction within the UAS which, due to either weaker binding affinities or perhaps nuclear matrix association,

are undetectable by either band shift or *in vitro* footprint techniques. *In vivo* footprinting might be a more efficient method for the detection of such interactions. As demonstrated for the human *PGK-1* promoter, *in vitro* binding techniques failed to reveal even half of the protein binding sites revealed by *in vivo* techniques (Figure 5, Chapter 1).

The UAS was originally restricted to the sequence between the Eco R1 and Stu 1 restriction enzyme sites. Both the Eco R1 - Alu 1 and Alu 1 - Stu 1 fragments were necessary for enhancement activity. It is curious, therefore, that both R1 and R2 are located in the proximal region of the UAS, within the Alu 1 - Stu 1 sequence, and no protein:DNA binding sites were found in the distal region, within the Eco R1 - Alu 1 sequence. One model which would explain how the distal region is involved in transcriptional enhancement has the distal DNA interacting with the protein already bound to the sites in the proximal region. Repression of transcription would therefore occur if protein interacted with DNA in the distal region and inhibited the interaction of the DNA with the proximal region binding proteins, or if the binding protein in the proximal region changed, and prevented the interaction with the distal region DNA.

Mutation of the R2 binding site, which resulted in a change in the protein-binding interaction, did repress transcription levels. It is unknown, however, whether this mutation had any effect on the ability of the distal region to interact with the proximal binding activities. The fact that the reduction in gene expression was only observed in stably transfected DNA suggests that DNA conformation is important to enhancement activity. Therefore, it could be hypothesized that the mutation at R2 resulted in a conformational change of the stably integrated DNA, such as the change to a DNA bend

angle by factors bound to R2, which prevented the upstream sequence from coming in contact with the proximal binding activites. The ability of the R2 binding activity to bend DNA could be examined using the protocol described by Giese et al (1992).

The fact that the R1 binding activity differs in P19 and HeLa cells suggests that R1 might be important in tissue-specific regulation of transcription. According to the model, to be discussed in Chapter 5, R1 binding activity may be important for the interaction of either distal DNA or distal DNA binding factors in the regulation of gene expression.

CHAPTER 4

The UAS in Differentiated Cells

4.1 Introduction

It was established in the previous chapter that the pluripotent, undifferentiated P19 cells contain nuclear factors which interact with the UAS of the murine *Pgk-1* gene. By band shift and footprint analyses these interactions were limited to two regions, designated R1 and R2, within the more proximal half of the UAS, relative to transcription initiation. Interaction of factors with R1 was apparently tissue-specific, and in P19 cells, was different from the interaction at R2. The R2 site was shown to be important for strong promoter activity.

This chapter deals with experiments which were carried out to (1) demonstrate the developmental transcriptional regulation of Pgk-1 in the P19 system, and (2) show the involvement of the UAS in this regulation. The first section concerns the examination of the effect of retinoic acid (RA) on gene expression, in both transient and stable

transfections. The next section examines the effect of RA on the binding of transcription factors to the UAS. Following this is a test of the specificity of the binding interaction of one of the RA-induced binding activities. Finally, the interactions of UAS binding nuclear factors, from both undifferentiated and differentiated cell extracts, are examined by competition band shift studies.

4.2 Results

4.2.1 Expression from the UAS containing promoter is down-regulated in differentiated P19 cells

Cells were transiently transfected with 13 μg of either test plasmid pET-lacZ (500 bp Eco-Taq region of the *Pgk-1* promoter, driving lacZ) or pST-lacZ (200 bp Stu-Taq region of the *Pgk-1* promoter, driving lacZ), and 2 μg of control plasmid pKJ17(*Pgk*-CAT). Cells were stably transfected with 7 μg of either test plasmid, 2 μg of control plasmid RSV-CAT, 5 μg of B17 and 2 μg of the selectable plasmid *Pgk*-puromycin. In the transient transfection, cells were exposed to RA (1) just prior to transfection, in which case the cells were exposed to RA for a total of two days before harvesting, (2) one day prior to transfection, for cells exposed to RA for a total of three days, (3) two days prior to transfection, for cells exposed to RA for a total of four days, etc. In the stable transfections, the cells were transfected on Day 1, passaged and placed in selection medium the following day and grown for a total of eight days. Colonies were pooled,

and the cells re-plated in the selection medium, now containing RA, for a further six, eight or ten days. Transient and stable transfection experiments were performed in duplicate.

As seen in Figure 31A, the activity of the UAS containing promoter in transiently transfected P19 cells was at least 10-fold greater than the level of activity from the promoter lacking the UAS. Exposure of the cells to RA for two days resulted in a two-fold decrease in activity from the promoter with the UAS. This activity continued to decrease until around seven days after RA-treatment, when the UAS apparently no longer contributed to expression, which was totally driven by the core promoter.

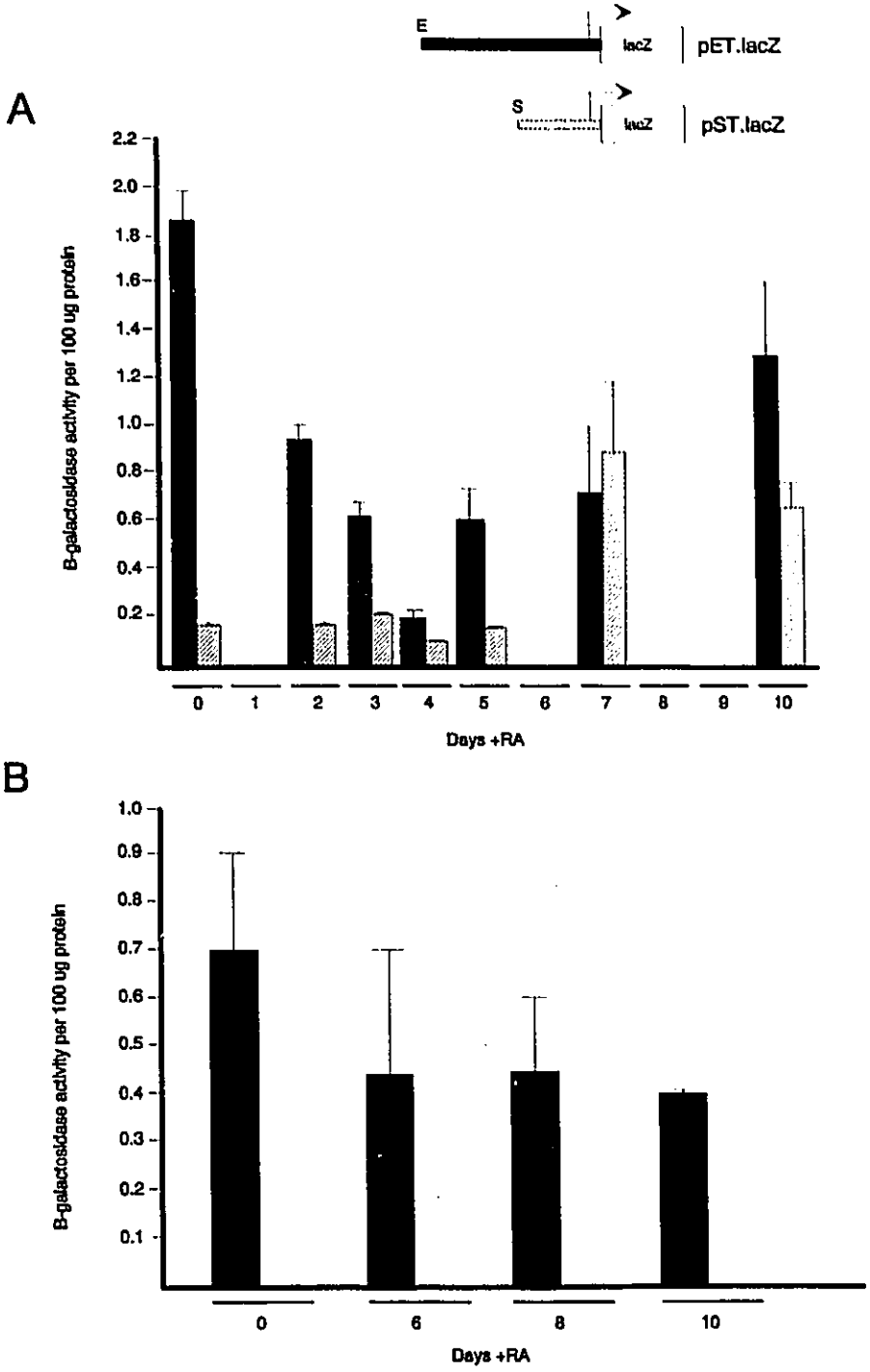
In the stable transfections (Figure 31B), the effect of RA-treatment on expression driven by the UAS containing promoter was not observed as quickly as in the transient transfections. After six days of RA-exposure, at a time when transiently transfected UAS containing promoter activity was essentially nil, activity of the UAS containing promoter in the stable transfections was still high. By Day 10 of RA exposure, activity relative to the before RA-treatment levels was comparable between the stable and transient transfections (approximately 60% activity). In the transient transfections, some of this activity was contributed by the core promoter, whereas in the stable transfections no core promoter activity was detected. As fewer copies of stably transfected DNA are present than transiently transfected DNA, the protein assay may simply not have had the sensitivity to detect very low levels of the reporter protein being expressed from the stably integrated core promoter.

Therefore, RA induces a down-regulation of UAS containing promoter driven

FIGURE 31: UAS activity in RA-treated P19 cells

(A) Transient transfection of plasmid containing lacZ driven either by the pgk-1 core promoter to -121 (pST.lacZ) or the pgk-1 promoter including the UAS to -424 (pET.lacZ). Cells were treated with RA, cultured for the required number of days, then transfected. Cells were harvested 48 hours after transfection. Days +RA includes those 48 hours. 13 μg of test plasmid were co-transfected with 2 μg of pKJ17 to measure transfection efficiency.

(B) Stable transfection of the same two test plasmids outlined above. Cells were transfected, and 24 hours later selected in puromycin for stable transformants. Colonies were grown for about eight days. Pools were then subcultured, and grown for the period of time indicated in RA before harvesting. 7 μg of test plasmid were co-transfected with 2 μg of RSV-CAT to measure transfection efficiency, 5 μg of B17 to increase the number of transformants, and 2 μg of pgk-puromycin for selection. Both transfection experiments were performed in duplicate. Error bars represent the standard error of the mean. P=0.05.


A
E
lacZ
pET.lacZ
S
lacZ
pST.lacZ
B-galactosidase activity per 100 ug protein
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
1
2
3
4
5
6
7
8
9
10
Days +RA
B
B-galactosidase activity per 100 ug protein
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
6
8
10
Days +RA

gene expression, and has a more immediate effect on transiently transfected DNA than on stably transfected DNA. As transiently transfected DNA is not incorporated into the chromatin, perhaps it is more accessible to the factors which are responsible for altering the expression levels. As there is a mixed population of cell types in the culture, including neurons, astrocytes and fibroblast-like cells, it is conceivable that a more homogeneous population would produce a greater reduction is gene expression.

Two important pieces of information can be gathered from these transfection experiments. One is that RA-induced differentiation leads to reduced UAS aided expression. The other is that the UAS plays a significant role in elevating expression from the core promoter when integrated into the chromatin. This suggests that the UAS may function as an enhancer of gene expression by altering chromatin structure.

4.2.2 Nuclear extract from differentiated P19 cells contains novel DNA binding protein

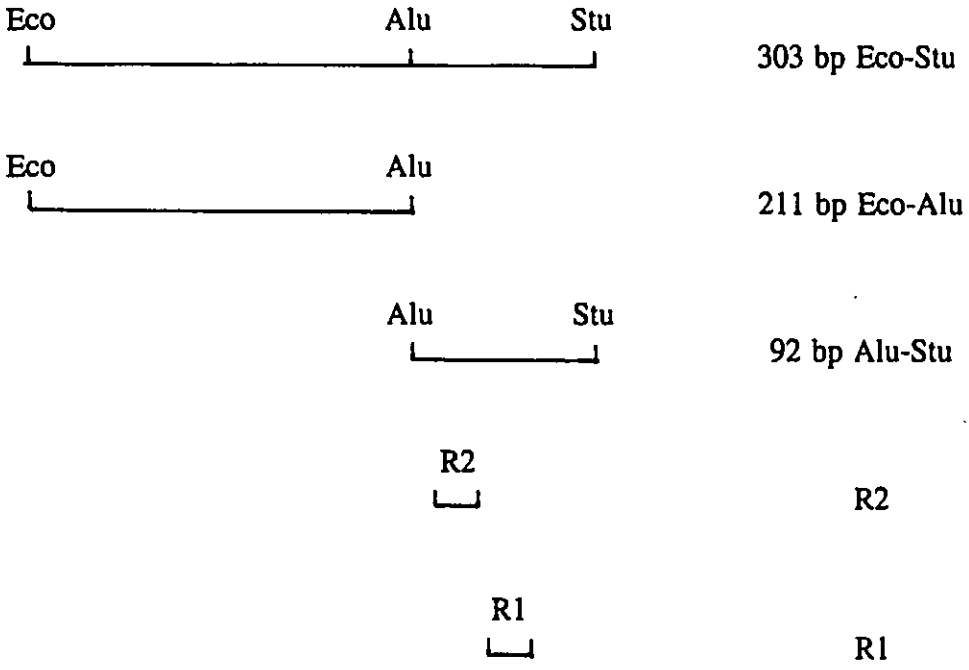
A comparison of binding activity between nuclear extract from P19 cells and nuclear extract from RA-treated P19 cells was performed in order to determine whether the RA-induced reduction in UAS enhanced gene expression was accompanied by an alteration in UAS DNA-binding interactions.

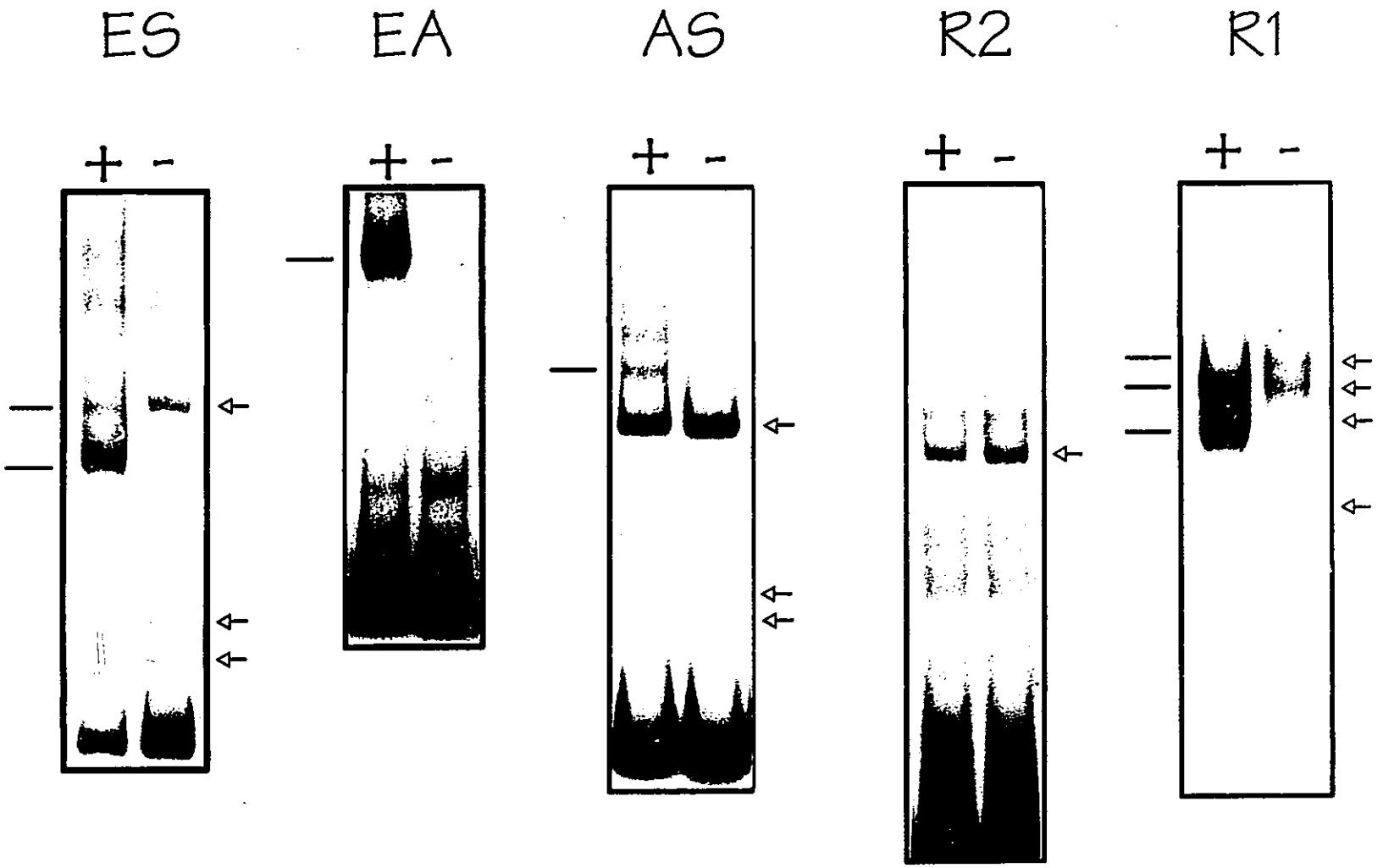
P19 cells were exposed to RA continuously for eight days at which time nuclear extract was purified. Band shift assays were performed using various restriction fragments of the UAS, comparing the RA-treated and untreated P19 nuclear extracts.

As demonstrated in Figure 32, binding to the 303 bp Eco-Stu UAS fragment

FIGURE 32: Nuclear extracts from differentiated P19 cells contain novel DNA binding protein

Band shift assays were performed using various restriction fragments of the UAS, as outlined in Chapter 2. Nuclear extract from P19 cells or P19 cells exposed to RA for eight days was used. The fragments tested were the 303 bp Eco-Stu (ES), the distal 211 bp Eco-Alu (EA), the proximal 92 bp Alu-Stu (AS), and the 30 bp R2 and R1 oligonucleotides. DNA was incubated with 20 μg of P19 nuclear extract which had not been exposed to RA (-), or which had been exposed to RA (+). Arrows to the right of the boxes indicate binding complexes in the untreated P19 nuclear extract. Lines to the left of the boxes indicate where an alteration in band mobility has occurred following exposure of the cells to RA. Assays were performed a minimum of five times.




ES
EA
AS
R2
R1
+ -
+ -
+ -
+ -
+ -

resulted in a shift of the mobility of the upper retarded complex and in the disappearance of one of the lower retarded complexes, using RA-treated extract. The UAS was then divided by the Alu restriction enzyme into the distal 211 bp Eco-Alu fragment and the proximal 92 bp Alu-Stu fragment. Binding to both the distal and the proximal fragments was effected by RA-treatment of the cells. On the distal 211 bp Eco-Alu fragment, where no binding had ever been detected using untreated P19 nuclear extract, binding activity was now observed using the RA-treated extract. On the proximal 92 bp Alu-Stu fragment an additional retarded complex was detected using the RA-treated extract. On both the distal and proximal fragments, using the untreated extract, a few retarded bands were observed which had not previously been detected, likely due to the fact the band shift conditions were optimal. R1 and R2 are both subregions of the Alu-Stu proximal sequence. The intensity of two of the retarded R1 complexes increased significantly using RA-treated extract, and the most retarded complex present when using the untreated extract disappeared. No change in retarded complexes was observed on the R2 DNA.

Further band shift assays were conducted, using various restriction fragments, to narrow down the site of binding interaction of the novel binding activity associating with the distal region of the UAS in RA-treated P19 nuclear extract. It was ultimately determined that binding was restricted to a 111 bp region defined by the Sph 1 and Hha 1 restriction sites. This 111 bp region was designated R3.

In summary, exposure of P19 cells for eight days to RA affected the binding of nuclear factors to the proximal region of the UAS, at R1, and resulted in a novel binding activity associating with the distal region of the UAS, at R3. Therefore the RA-induced

decrease in transcription activity from the UAS containing promoter is accompanied by a change in the binding of transcription factors to the UAS.

4.2.3 R3 binding activity is distinct from proximal UAS binding activity

It has now been established that the UAS of the *Pgk-1* gene binds nuclear factors, and that these factors change in concert with a reduction in gene expression accompanying differentiation.

The R3 region of the UAS was identified as a protein binding site exclusive to proteins from differentiated cells. The 111 bp stretch of DNA comprising R3 contains two E-boxes and an inverted Ets consensus binding sequence. Given that the R1 and R3 sites share E-box and Ets consensus sequences, and that binding to the R1 site was also influenced by differentiation, it was decided to determine if the R3 binding activity also interacted with R1 DNA.

It had previously been determined that Fraction 18 of the NE-17 preparation of fractionated RA-treated P19 nuclear extract contains the R2 binding activity (Figure 17). It is also known that the R2 binding activity elutes from the sephacryl-300 column before WiF-1 (examined by Dr.St-Arnaud, and known to elute in Fractions 20-27). It was determined that the R3 binding activity eluted in Fraction 29 of the NE-23 preparation of fractionated RA-treated P19 nuclear extract (data not shown). In this preparation of nuclear extract the WiF-1 protein elutes in Fractions 16-27. WiF-1 is a 65 kD protein. Therefore, the proximal DNA binding activity which interacts at R2 has a DNA binding

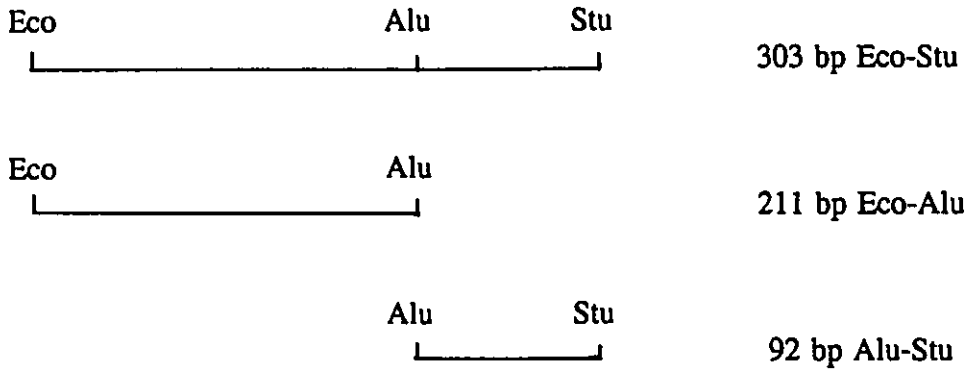
component with a molecular weight exceeding 65 kD, and the distal binding activity which interacts at R3 has a DNA binding component with a molecular weight less than 65 kD. Therefore, if protein from Fraction 29 were found to interact with the proximal DNA, it would suggest an interaction at R1 and not R2.

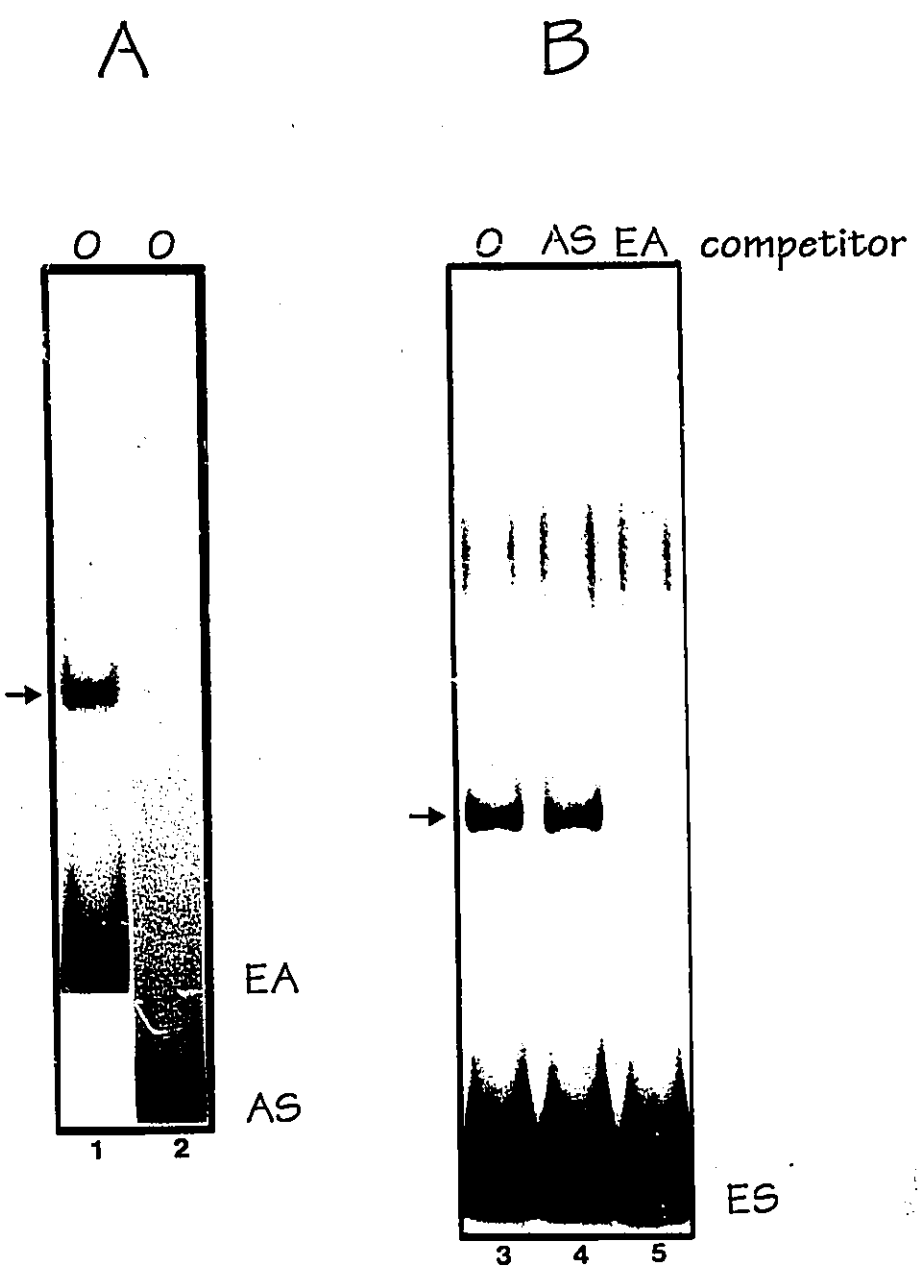
Two band shift experiments were performed. In the first experiment, the ability of proximal region DNA to interact with protein from the same column fraction which contained R3 binding activity was examined. Either the distal 211 bp Eco-Alu fragment or the proximal 92 bp Alu-Stu fragment was incubated with extract from Fraction 29 of the NE-23 preparation of fractionated RA-treated P19 nuclear extract (see Materials and Methods). In the second experiment, the ability of proximal DNA to compete for R3 binding activity was examined. The 303 bp Eco-Stu fragment was incubated with extract from Fraction 29, and unlabelled proximal 92 bp Alu-Stu DNA or distal 211 bp Eco-Alu DNA, as competitor.

The result of the first experiment is seen in Figure 33A. Only the distal 211 bp Eco-Alu fragment bound protein from Fraction 29. In the second experiment (Figure 33B), the proximal 92 bp Alu-Stu fragment was unable to compete with the 303 bp Eco-Stu fragment for protein from Fraction 29, whereas the distal 211 bp Eco-Alu fragment was able. These results demonstrate that no binding activity from Fraction 29 interacts with the proximal DNA, and suggest that the binding activity which interacts at R3 is not the same as the binding activity which interacts at R1 in differentiated cells.

FIGURE 33: R3 binding activity is distinct from proximal binding activity

Fraction 29 was used in band shift assays to determine specificity of binding. **(A)** The distal 211 bp Eco-Alu (EA) fragment (lane 1) and the proximal 92 bp Alu-Stu (AS) fragment (lane 2) were incubated with 5 μl of Fraction 29. **(B)** The 303 bp Eco-Stu (ES) fragment was incubated with 5 μl of Fraction 29, and 100-fold molar excess of competitor DNA. Lanes: 3, no competitor; 4, unlabelled (proximal) AS competitor DNA; 5, unlabelled (distal) EA competitor DNA. Arrows indicate retarded complexes. Assays were performed three times.




A
B
0 0
0 AS EA competitor
EA
AS
ES
1 2
3 4 5

4.2.4 R3 and R1 are key sites in the developmental transcriptional regulation of *Pgk-1*

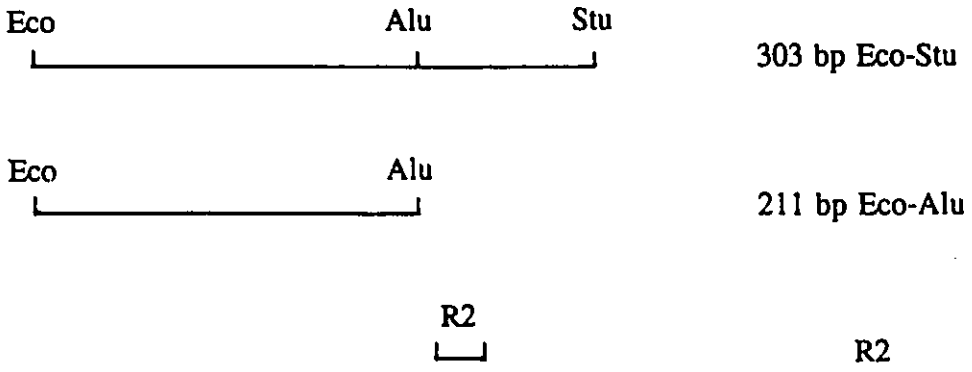
The binding of nuclear protein from undifferentiated and differentiated cell extracts, to the R2 oligonucleotide, was shown by band shift analysis to be identical (Figure 32). However, gene expression was enhanced in undifferentiated cells compared to differentiated cells (Figure 31), and the R2 region was found to be important for that enhancement (Figure 25). It was therefore decided to examine the binding at R2 in relation to the surrounding sequence, and the presence of other DNA binding factors. The result of the experiment described below demonstrates that the presence of R3 binding protein inhibits the interaction of protein at R2.

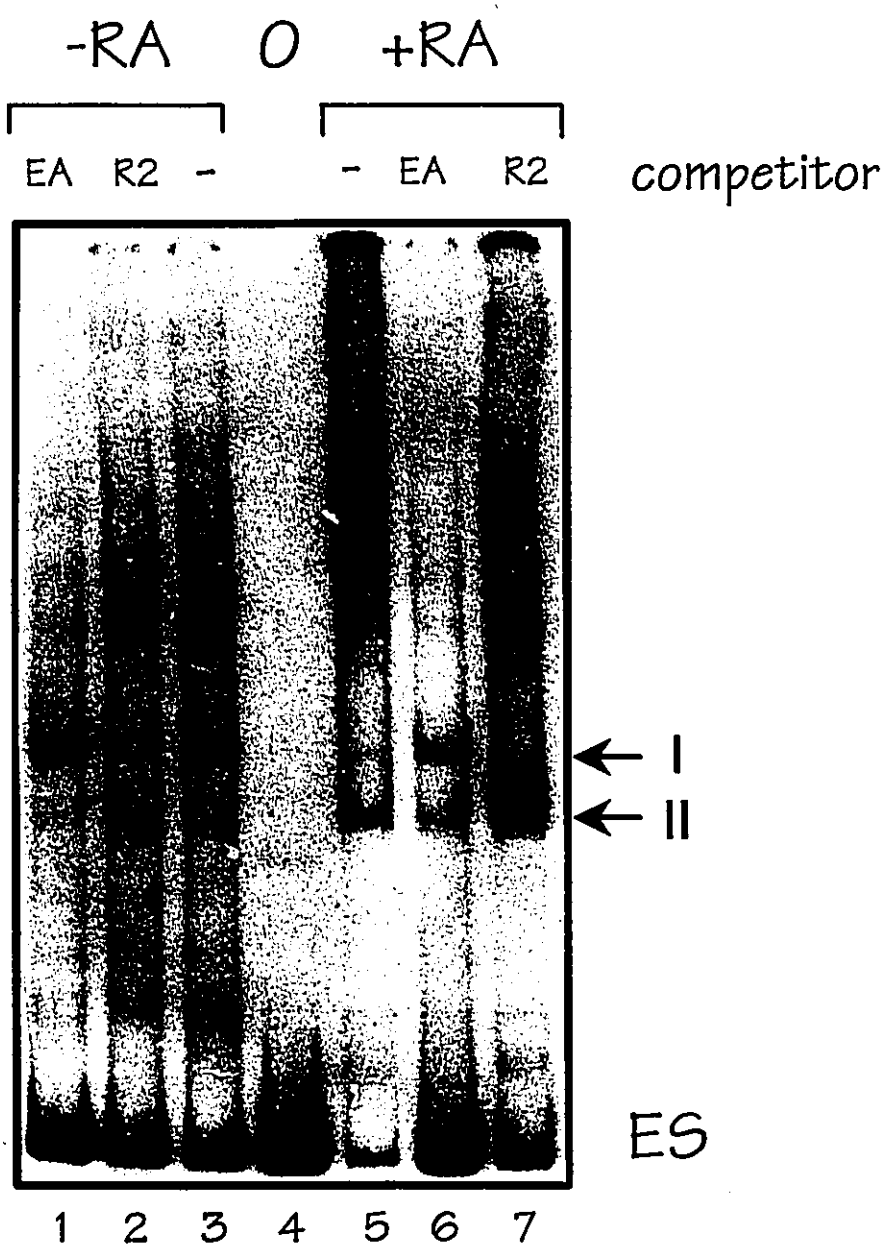
The 303 bp Eco-Stu fragment was incubated with P19 nuclear extract, or RA-treated P19 nuclear extract, in the presence of a 200-fold molar excess of unlabelled R2 oligonucleotide, or unlabelled distal 211 bp Eco-Alu DNA.

As seen in Figure 34, incubation of the labelled 303 bp Eco-Stu fragment with P19 nuclear extract in the presence of unlabelled R2 DNA resulted in the disappearance of the upper, most retarded band (Complex I). This band was uneffected by competition with the unlabelled 211 bp Eco-Alu DNA. Conversely, the lower, less retarded band (Complex II) was uneffected by competition with R2, but diminished upon competition with the distal 211 bp Eco-Alu DNA. Using RA-treated P19 nuclear extract resulted in an alteration of the banding pattern. Less DNA was bound in Complex I and more DNA was bound in Complex II (evidenced by the change in band intensities). Competition with the distal 211 bp Eco-Alu fragment under these conditions resulted in a diminution of

FIGURE 34: R3 binding activity interferes with proximal binding activities in RA-treated P19 cells

In a band shift assay, the labelled 303 bp Eco-Stu fragment was incubated with P19 nuclear extract, or P19 nuclear extract exposed for eight days to RA, in the presence of a 200-fold molar excess of unlabelled R2 oligonucleotide or unlabelled distal 211 bp Eco-Alu DNA. Lane 4 contains no extract, lanes 1-3 contain 20 μg of P19 nuclear extract, and lanes 5-7 contain 20 μg of RA-treated P19 nuclear extract. Arrows indicate the two retarded complexes. Assay was performed twice.




-RA
0
+RA
EA
R2
-
-
EA
R2
competitor
I
II
ES
1
2
3
4
5
6
7

Complex II and an increase in the amount of DNA bound in Complex I. Competition with R2 did not effect binding in Complex I.

This experiment demonstrated that the presence of R3 binding activity inhibited the interaction of protein at R2. The apparent contradiction between this result and the data presented in Figure 32 can be explained by the fact that in the band shift experiment presented in Figure 32, the R2 oligonucleotide was used, meaning that the R2 protein binding site was isolated from the surrounding UAS sequence and therefore unable to be influenced by other binding factors. In the experiment described above, these sequences were included, and the inhibition by R3 binding activity was observed. The ability to obtain a footprint at R2 using nuclear extract from RA-treated P19 cells was obviously due to the fact that fractionation of the extract separated the R3 and R2 binding activities.

This suggests that the binding of protein at R3 either (1) destabilizes the binding of protein at R2 by altering DNA configuration, or (2) prevents the binding of protein at R2 by steric hindrance. As the R2 and R3 binding regions are separated by a minimum of 54 bp, this would imply that some mechanism was functioning in the differentiated cells to bring the two regions into close enough proximity for steric hindrance to occur. To determine whether this mechanism involves R1, and the interaction of R1 and R3 binding activities (which might result in steric hindrance at R2 due to the close proximity of the R1 and R2 sites), the following experiment was performed. The ability of the R3 binding activity to inhibit the protein:DNA interaction at R2 was examined using undifferentiated P19 cell extract, as the R1 binding activity is different in differentiated and undifferentiated cells.

The 251 bp Sph-Stu fragment was incubated with either P19 nuclear extract, Fraction 29 or different quantities of Fraction 29 mixed with a constant quantity of nuclear extract. It was anticipated that if the binding activity in Fraction 29 inhibited factors present in the nuclear extract from interacting with the DNA, then a decrease in the nuclear extract retarded complex would be observed upon co-incubation of the DNA with Fraction 29.

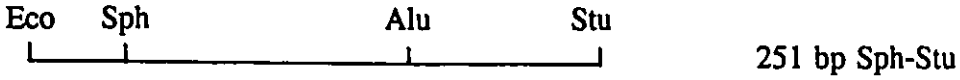
This was not observed. As seen in Figure 35, mixing Fraction 29 with untreated nuclear extract did not result in a decrease of the nuclear extract retarded complex.

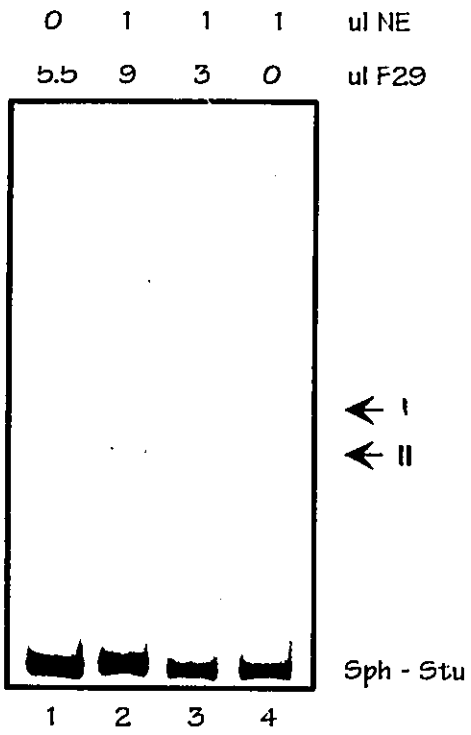
Therefore, the R3 binding activity in Fraction 29 was not able to interfere with, or prevent the binding of, the proximal binding activities present in undifferentiated P19 nuclear extract.

The only detectable difference in the binding of protein to the proximal region of the UAS between differentiated and undifferentiated P19 cells was at R1 (Figure 32). Therefore, the ability of the R3 binding activity to interfere with the binding of protein to R2 was apparently influenced by the protein interaction at R1. However, the alteration of the binding interaction at R1 following RA-treatment could either (1) stimulate the association of the R3 and R1 binding activities, and incidentally hinder the interaction of factors at R2, or (2) weaken the DNA-binding interaction at R2 and make it vulnerable to displacement by the R3 binding activity. Comparing the affinities of the R2 binding activities from differentiated and undifferentiated cell extracts, in band shift assays using a DNA fragment containing only proximal sequence, would be one means of determining how the alteration of the binding interaction at R1 affects the R2 protein binding

FIGURE 35: R3 binding activity does not interfere with proximal binding activities

In a band shift assay, the labelled 251 bp Sph-Stu fragment was incubated with P19 nuclear extract and/or Fraction 29, NE-23. Lanes: 1, 5.5 μl of Fraction 29; 2, 9.0 μl of Fraction 29 mixed with 12 μg of nuclear extract; 3, 3.0 μl of Fraction 29 mixed with 12 μg of nuclear extract; 4, 12 μg nuclear extract alone. Arrows indicate the two retarded complexes. Assay was performed twice.




0 1 1 1 ul NE
5.5 9 3 0 ul F29
I
II
Sph - Stu
1 2 3 4

interaction.

In summary, the results of the two band shift experiments described above suggest that both R1 and R3 are important protein-binding sites for the developmental regulation of *Pgk-1* transcription.

4.2.5 R2 is not involved in the repression of gene expression in differentiated cells

To confirm the lack of involvement of the R2 site in the transcription of PGK-1 in differentiated cells, the effect of the R2 mutation on expression in RA-treated P19 cells was examined.

Both transient and stable transfections were performed, in duplicate. P19 cells used in the transient transfections were exposed to RA for 8 days prior to transfection. Pools of clones, of the same stably transfected P19 cells that were used for the test of the R2 mutation in untreated P19 cells, were exposed to RA for 8 days.

Comparing expression levels of the enzyme driven by the promoter carrying the R2 site with the promoter carrying the mutated R2 site, in RA-treated P19 cells, no significant change was observed in either transient (Figure 36A) or stable transfections (Figure 36C).

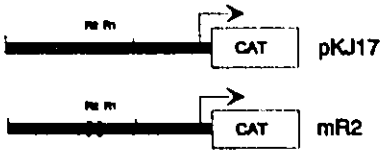
These results confirm that the R2 binding site is not important for expression in RA-treated P19 cells.

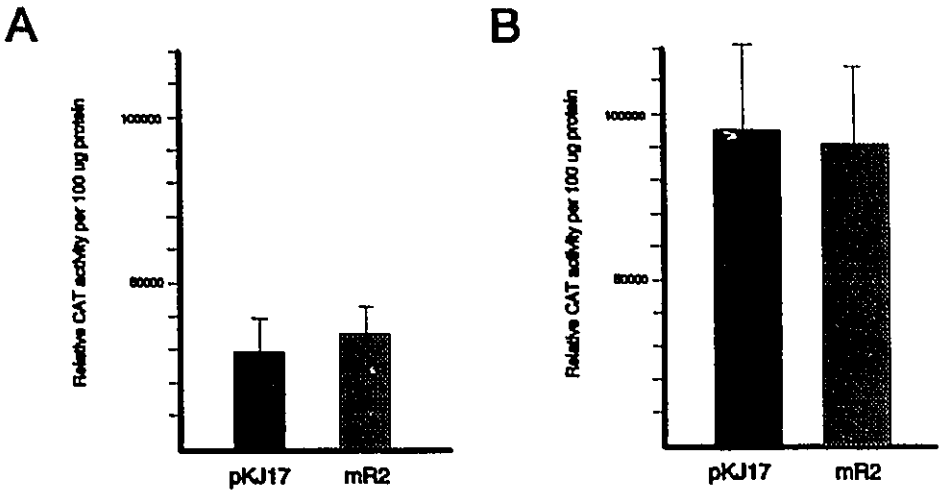
FIGURE 36: Involvement of R2 site in RA-treated P19 cells

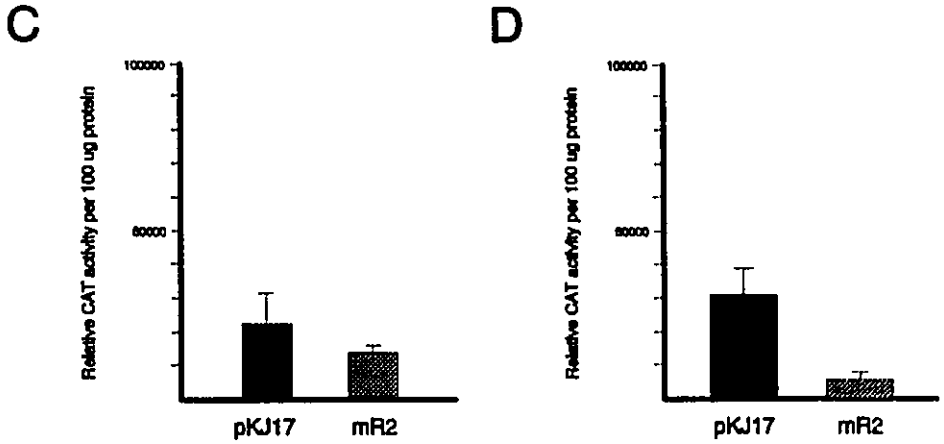
The results presented in graphs **(A)** and **(C)** were obtained using P19 cells exposed to RA for eight days. The results in graphs **(B)** and **(D)** were obtained using undifferentiated P19 cells, and were already presented in Figure 25 but are included here for comparison purposes.

(A), (B) Transient transfection of either the wild-type pgk-1 UAS in the pKJ17 plasmid, or the pgk-1 UAS carrying the R2 mutation in the mR2 plasmid. 5 μg of test plasmid were co-transfected with 5 μg of the control plasmid pgk-lacZ.

(C), (D) Stable transfection of the same test plasmids. 5 μg of test plasmid were co-transfected with 5 μg of control plasmid, 2 μg of B17 and 2 μg of pgk-puromycin. Both transfection experiments were performed in duplicate. Error bars represent the standard error of the mean. P=0.05.


CAT
pKJ17
CAT
mR2


A
B
Relative CAT activity per 100 ug protein
100000
80000
pKJ17
mR2


C
D
Relative CAT activity per 100 ug protein
100000
80000
pKJ17
mR2

4.3 Conclusions and Discussion

Once P19 cells begin to differentiate, gene expression from the murine *Pgk-1* promoter is repressed, and the interaction of transcription factors with the UAS changes. This suggests that the UAS binding proteins are involved in the transcription repression associated with differentiation.

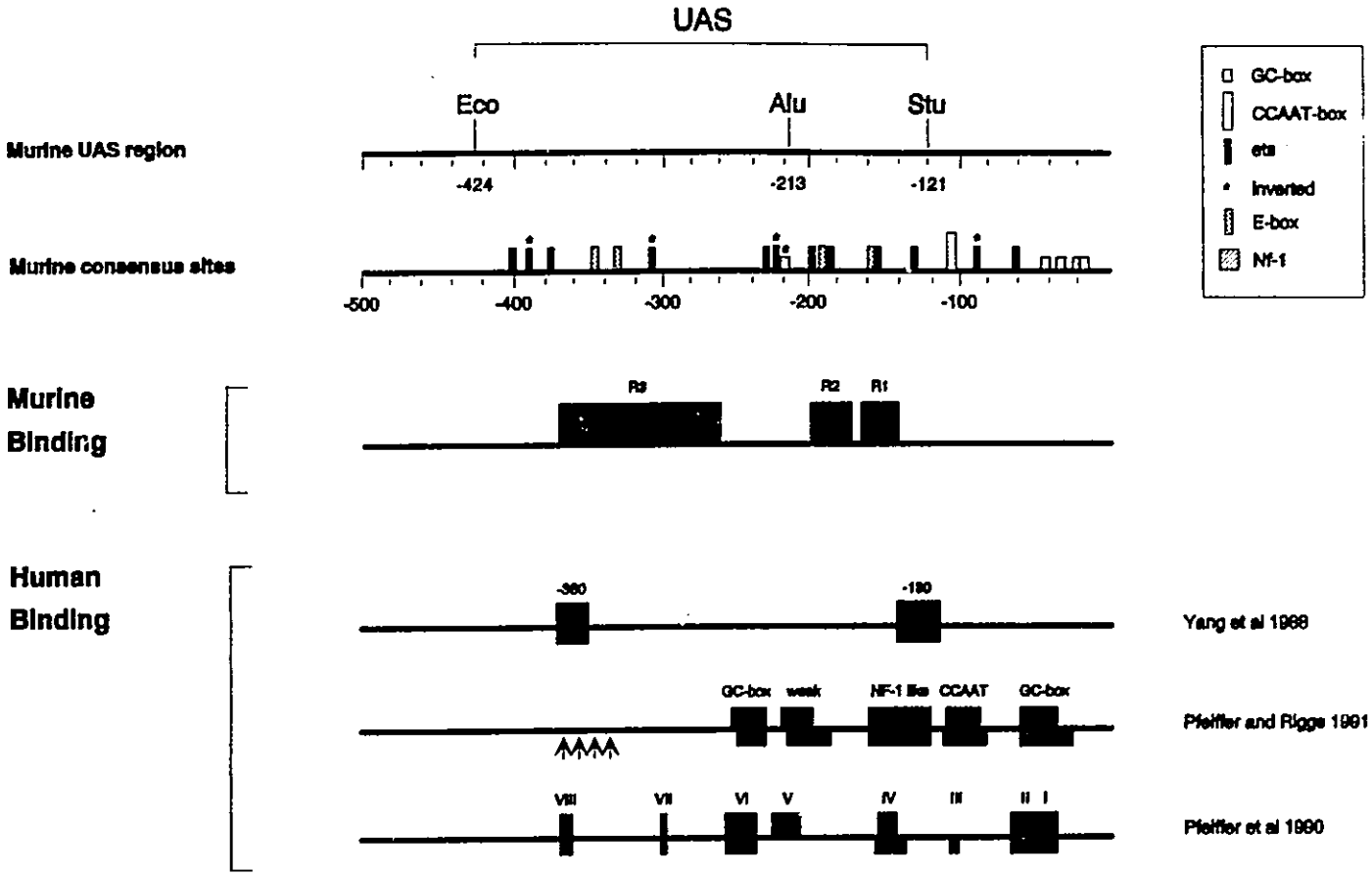
In the differentiated cells, protein interaction was detected at R1 and R3 but not R2, and the results of expression studies then supported the hypothesis that R2 was therefore not important for expression in RA-treated P19 cells. The observed interaction of factors with R1 by band shift analysis was found to differ between the undifferentiated and differentiated cells (Figure 32). As this interaction was already found to differ between P19 and HeLa cells (Figure 30), it suggests that R1 is a pivotal protein binding site in the tissue-specific and developmental regulation of PGK-1 gene expression.

A model which explains how the interaction of transcription factors from differentiated cells represses expression of the *Pgk-1* gene would have the activator protein(s) which associates with R2 in undifferentiated cells prevented from binding DNA by the presence of a repressor protein at R3, and an alteration in protein binding at R1. DNA in the distal region would be prevented from forming the same associations with the proximal region as in undifferentiated cells, and the resulting change would be unfavourable for high levels of transcription initiation.

R1 and R3 overlap with known sites of protein interaction on the human *PGK-1* promoter (Figure 37). The cells used in the human studies were either lymphocytes or

FIGURE 37: Murine UAS protein binding sites in relation to the known human protein binding sites

The R1, R2 and R3 sites of protein:DNA interaction were identified in this thesis. TIN-1 was discussed in Chapter 1, and is a tissue-specific repressor protein of PGK-1 in rat. The known sites of protein:DNA interaction on the human *PGK-1* promoter are included for comparison purposes. Consensus binding sequences on the murine *Pgk-1* promoter are included for reference.


UAS
Eco
Alu
Stu
Murine UAS region
-424
-213
-121
Murine consensus sites
-500
-400
-300
-200
-100
GC-box
CCAAT-box
ets
Inverted
E-box
Nf-1
Murine
Binding
R3
R2
R1
Human
Binding
-360
-180
Yang et al 1988
GC-box
weak
NF-1 like
CCAAT
GC-box
Pfeiffer and Rigge 1991
VIII
VII
VI
V
IV
III
II
I
Pfeiffer et al 1990

HeLa, both differentiated cell types. It was therefore of interest to note that the human homologue of the murine R2 site was not detected using any of the protein binding detection techniques, save *in vivo* DNase 1 footprinting, where a weak interaction was observed (Pfeiffer and Riggs, 1991). This observation correlates with the findings reported in this chapter, which noted that the presence of a distal binding activity in differentiated P19 cells inhibited binding in the R2 region, and suggests that the same mechanism might be operating in the human cells. The differental binding activity associated with the murine promoter is, therefore, not murine specific.

Both R3 and the distal protein binding site in the human promoter contain an Ets consensus binding sequence. The repressive mechanism which operates in the differentiated P19 cells might, therefore, involve Ets factors. Two factors which are known repressors of PGK-1 transcription, PEA3 and TIN-1, are Ets proteins (Hassel, unpublished; Goto et al 1991). It is unlikely, however, that the PEA3 protein is involved in the repression of transcription in either the HeLa cells or the differentiated P19 cells, as the protein has a highly restricted expression pattern (being readily detectable only in epididymis, brain and undifferentiated P19 cells) and its mRNA has been shown to be down-regulated during the retinoic acid-induced differentiation of P19 cells (Xin et al 1992). Also, TIN-1, which has been shown to repress transcription of the rat *Pgk-1* gene during spermatogenesis (Chapter 1), binds to a palindromic Ets sequence in rat which is conserved in mouse and is located just 5' to the Alu restriction site within the UAS, but is therefore not within the R3 region (Goto et al 1991).

CHAPTER 5

Discussion

Three sites of protein:DNA interaction, R1, R2, and R3, have been identified using *in vitro* binding techniques on the murine *Pgk-1* UAS. R3 interacts with protein from RA-treated P19 cells, R1 interacts with protein from both untreated and RA-treated P19 cells, and R2 interacts with protein from both untreated and RA-treated P19 cells but is unable to bind protein from RA-treated P19 cells in the presence of the R3 binding activity. It was determined by binding analyses of fractionated nuclear extracts that the R2 and R3 binding activities are different. It was also determined that transcription activity from the *Pgk-1* promoter is reduced in RA-treated P19 cells. It can therefore be concluded that multiple factors, which are developmentally regulated, interact with the UAS of the *Pgk-1* gene in a manner which is correlated to gene expression.

Model

One explanation to account for how the interaction of transcription factors with the UAS of the murine *Pgk-1* promoter modulates expression is diagramatically presented in Figure 38. It is based on the importance of the R1 site in PGK-1 transcription regulation.

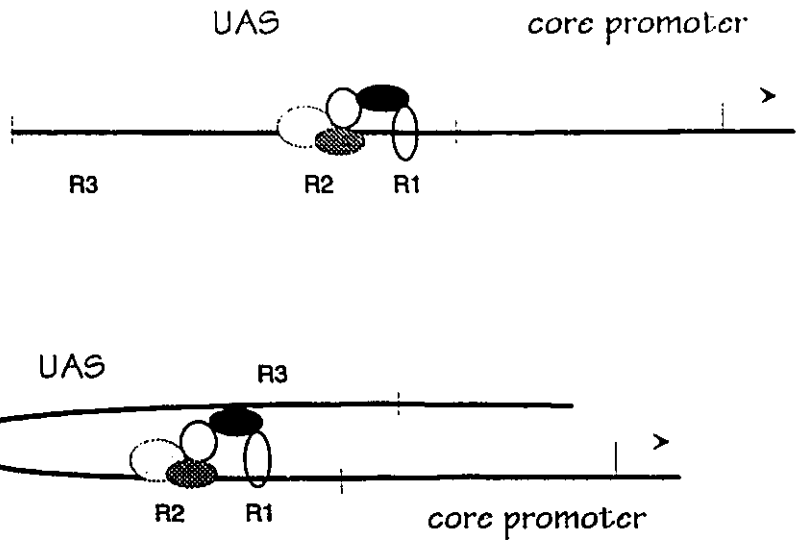
The model depicts DNA in the distal region of the UAS interacting with the proximal binding activities in undifferentiated P19 cells. It is hypothesized that the R1 binding activity interacts with the distal DNA, and the R2 binding activity. Interaction of the R1 and R2 binding activities is hypothesized to stabilize the low affinity interaction at R1. Following differentiation, the binding activity at R1 is altered, and a new binding activity associates with R3. It is hypothesized that the R3 and the R1 binding activities are able to interact, and the interaction results in the loss of binding at R2 due to steric hindrance. The conformation of the DNA is then unfavourable for high levels of gene expression.

The model is supported by the following observations. The distal region of the UAS is known to be required for elevated gene expression in undifferentiated P19 cells, and the R1 and R2 sites were found to bind protein. The R2 binding activity is depicted as a protein complex because the mutation which disrupted so much of the sequence did not completely abolish binding activity, which is consequently hypothesized to involve at least two DNA-binding components. R1 is depicted as a protein complex because in the binding experiments using the heparin-agarose fractionated nuclear extract, one retarded complex was observed, which suggests that a single DNA-binding factor is present, and that the complex banding pattern obtained using unfractionated nuclear

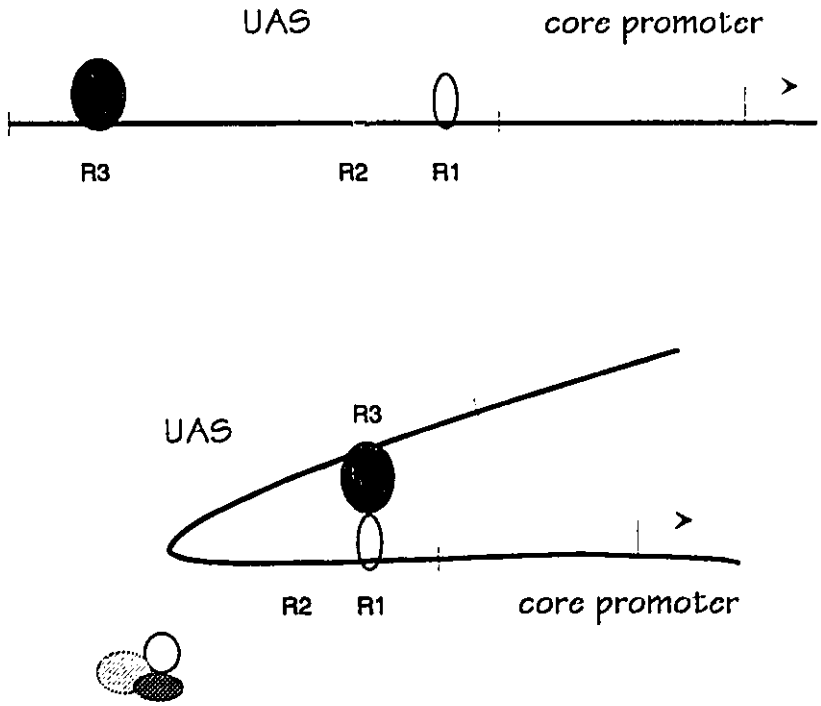
FIGURE 38: Model

A model of the interaction of P19 cell transcription factors with the UAS of the murine pgk-1 gene. For details see text.

-RA


UAS
core promoter
R3
R2
R1
UAS
R3
R2
R1
core promoter

+RA


UAS
core promoter
R3
R2
R1
UAS
R3
R2
R1
core promoter

extract indicates the presence of additional non-DNA binding proteins associating with R1 (Figure 28). In differentiated cells, the alteration in R1 binding activity is depicted as the loss of a member of the protein complex, and a new association with the R3 binding activity.

The presence of the new R3 binding activity in RA-treated cells, and the change in the R1 binding activity, were observed by band shift analysis using extract from RA-treated P19 cells. The loss of binding at R2 in RA-treated P19 cells was deduced from the band shift assay in which competition of R3 binding activity from RA-treated cells, by distal UAS DNA, resulted in complex formation associated with R2. This demonstrated that binding activity is not associated with R2 in the presence of R3 binding activity.

The pivotal role of R1 in the developmental transcriptional regulation of *Pgk-1* could be demonstrated by mutating the site, and examining the effect of the mutation on gene expression in both untreated and RA-treated P19 cells. One would expect that gene expression would be reduced in both cases, for although R1 is hypothesized to be involved in the reduction of expression in RA-treated cells, it is also hypothesized to be necessary for the elevation of gene expression, associated with untreated P19 cells.

An experiment which could be performed to examine the association of R1 and R3 binding activities is a competition band shift analysis, involving RA-treated nuclear extract, R1 DNA, and distal region DNA. When labelled R1 DNA is incubated with RA-treated nuclear extract in the presence of unlabelled R3 competitor DNA, then if R3 binding activity was interacting with the R1 binding activity, it would be competed away

and the banding pattern after competition would resemble the banding pattern using untreated P19 nuclear extract. Another way of examining the same question would be to incubate Fraction 29 and nuclear extract from untreated P19 cells with R1 DNA. If the resulting banding pattern in a band shift assay resembled the banding pattern observed when RA-treated extracts were used, then it would suggest that in the RA-treated cells, the R3 binding activity is interacting with the R1 binding activity.

It is not known, because the studies were not carried that far, if the reduced expression of PGK-1 is maintained in the retinoic acid-induced cells, or whether it is a temporary phenomenon associated with a reduced requirement for phosphoglycerate kinase during the immediate reprogramming of the cells. Obviously not all differentiated cells have reduced levels of phosphoglycerate kinase, but all may experience a reduction in transcription activity during the initial stages of differentiation. It would be interesting to (1) extend the analysis of PGK-1 expression in retinoic acid-induced differentiation over a longer time frame, and perhaps in a more homogeneous population of cells, and to (2) parallel the studies in DMSO-treated P19 cells, which are induced to form muscle cells, which can have requirements for high levels of PGK-1. This would help us to understand whether the observed repression in *Pgk-1* gene expression is restricted to neuronal differentiation or whether it is a general phenomenon which cells experience during the initial stages of differentiation. If it is a general phenomenon, then transcriptional mechanisms must also operate to re-activate the gene in a tissue-specific manner. This opens up possibilities for another line of investigation.

APPENDIX I

Southwestern Analysis

Introduction

Southwestern blotting (Vinson et al 1988) is the DNA probing of protein immobilized on a matrix. When the protein is part of a mixture, for instance a nuclear extract, and is electrophoresed through a denaturing SDS gel prior to transfer, this technique can be used to determine the number and molecular weight of proteins in the mixture which bind a particular sequence of DNA. The technique will not yield results if the protein requires subunit association in order to bind DNA. It may therefore be a useful exercise to perform southwestern analysis prior to screening an expression library, where individual recombinant peptides are produced by each phage.

The results are presented in an appendix due to the fact that consistent data was hard to obtain. The lines in Figure 39 indicate bands which were observed in a minimum of two experiments.

Materials and Methods

Heparin-agarose Column 2. 200 μl (approximately 4.6 mg) of P19 nuclear extract from a single preparation were used. This was diluted with 300 μl of nuclear extract Buffer D (see above) before column application. The extract was left to sit 10 min before elution. 500 μl aliquots were collected with the same elution strategy used for Column 1 above. 18 fractions were collected and concentrated directly, without initial dialysis. After the first centrifugation to concentrate each fraction, 1.5 ml of Buffer D were added to the fractions, which were then re-centrifuged for an additional two hours at 5000 g. The concentrators containing fractions 9-14 were filled with an additional 1 ml of Buffer D and centrifuged another two hours. Fractions 12 and 13 were centrifuged for three hours. Protein concentrations were determined and fractions 5, 9, 13 and 16 were found to contain the highest levels of protein.

Column 3. 110 μl (2 mg) of P19 nuclear extract from a single preparation were used. 250 μl aliquots were collected. Salt elutions were slightly different than above, being 0.1 M, 0.2 M, 0.3 M, 0.4 M and 0.5 M KCl in Buffer D. The extract was not diluted before column application and consequently the first four fractions eluted over a period of 2.5 hours. Samples were dialysed overnight against Buffer D. Protein concentrations were determined and fractions 3, 8, 12, 16 and 19 contained the highest quantities of protein. The fractions were not concentrated. 5 μl were electrophoresed on a 7.5% SDS polyacrylamide gel.

Extract was electrophoresed through 7.5% SDS gels made with a 38:1 acrylamide:bisacrylamide mixture. Boiling of protein samples prior to electrophoresis did

not affect binding, therefore was not routinely performed. Gels were transferred to HyBond-C nitrocellulose membranes at 80 volts for one hour. After transfer, the membranes were air-dried for 30 min prior to being blocked in a buffer containing 50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT and 5% dried milk powder, for one hour at room temperature. Fresh buffer, minus the milk, was added to the membrane at a volume of 0.1 ml/cm_2, containing 10 μg/ml of poly(dI-dC). After one hour at room temperature the membrane was exposed to fresh buffer, minus the poly(dI-dC), to which labelled probe had been added at a concentration of 1-2 x 10^6 cpm/ml. The binding reaction was left for two hours at room temperature. The membranes were rinsed in the original blocking buffer, minus the milk, five or six times over 30 min. They were sealed in plastic and exposed to X-ray film with a screen, overnight, at -70°C.

In experiment #1, 10 μg of P19 nuclear extract, 10 μg of cytoplasmic nuclear-free cell extract, and 3 μl of Fraction 7 from the heparin-agarose columns #1 and #3 were examined for DNA binding factors. The membrane was probed with the 158 bp Bam-Eco fragment of the pLSe140 plasmid, containing 136 bp of the proximal UAS between the Hha 1 and Stu 1 restriction sites. In experiment #2, either 18 or 36 μg of a different preparation of P19 nuclear extract was tested and the membrane was probed with the 302 bp Bam-Eco fragment of the pLSe140x2 plasmid, containing two copies of the 136 bp Hha 1 - Stu 1 restriction fragment in a head-to-tail orientation. In experiment #3, 12 μg of another preparation of P19 nuclear extract, and 5 μl of Fractions 7, 8 and 9 from the heparin-agarose column #2 were tested. The probe used was the 158 bp Bam-Eco fragment of pLSe140, as in the first experiment. In experiment #4, 36 μg of the same

P19 nuclear extract used in experiment #2, 72 μg of a new preparation of P19 nuclear extract, and 7 μg of Fraction 17 from the NE-17 nuclear extract fractionation of P19 cells exposed for four days to retinoic acid were tested. The probe used was again the 158 bp Bam-Eco fragment of plasmid pLSe140.

Results

In the first experiment (Figure 39A), the nuclear extract (lane 4) yielded multiple signals, including a dark one around 120 kD, faint ones at around 90 and 70 kD, and a cluster of dark ones between 40 and 50 kD. The cytoplasmic nuclear-free cell extract (lane 3) produced a faint signal around 90 kD. Fraction 7 from column #1 (lane 2) resulted in a single signal at around 120 kD, whereas Fraction 7 from column #3 (lane 1) resulted in two dark signals around 50 kD, in addition to the 120 kD signal.

In the second experiment (Figure 39B), 36 μg of a different P19 nuclear extract preparation (lane 2), probed with the 302 bp fragment, resulted in a sharp signal at around 250 kD, another at around 100 kD, and a number of much fainter, hazy bands, including one around 70 kD. The 250 kD band was not observed in the lane containing 18 μg of protein (lane 1).

In the third experiment (Figure 39C), binding of nuclear extract, from yet another preparation, and probed with the 158 bp fragment, was very fuzzy. The combined Fractions 7 and 8 of the heparin-agarose column #2 yielded no signal. However, in Fraction 9 of column #2 a band at around 110 kD was just detectable.

In the fourth experiment (Figure 39D), the lane containing 72 μg of nuclear

FIGURE 39: Southwestern analysis of P19 nuclear extract

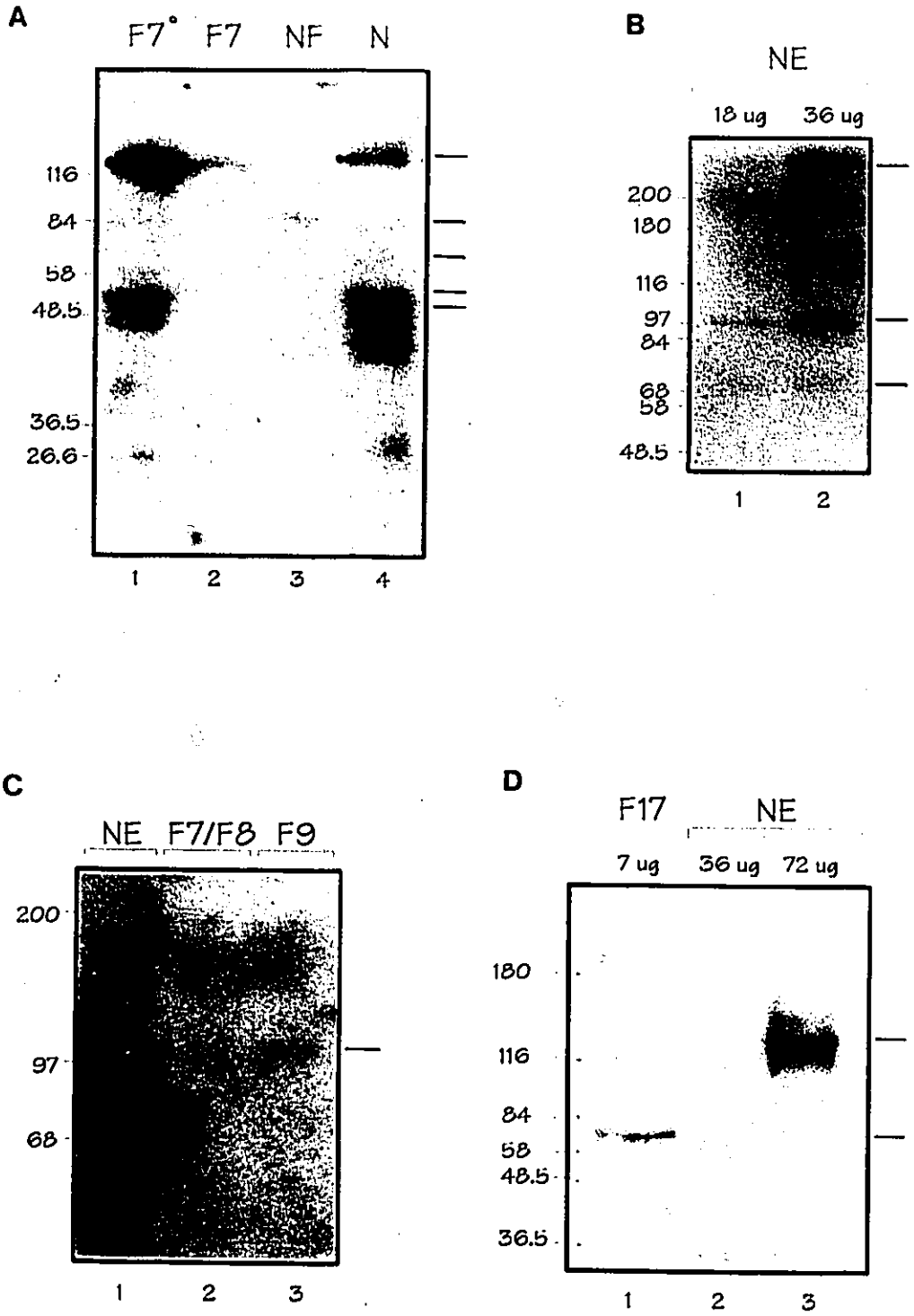
The results of four individual experiments are presented. Lines at the side of the boxes indicate bands which are discussed in the text.

(A) Extracts were probed with the 158 bp Bam-Eco fragment of the pLSe140 plasmid, containing 136 bp of the proximal region of the UAS, between the Hha 1 and Stu 1 restriction sites. Lanes: 1, 3 μl of Fraction 7 from the heparin-agarose column #3 (F7°); 2, 3 μl of Fraction 7 from the heparin-agarose column #1 (F7); 3, 10 μg of cytoplasmic nuclear-free cell extract (NF); 4, 10 μg of P19 nuclear extract (N).

(B) Extract was probed with the 302 bp Bam-Eco fragment of the pLSe140x2 plasmid, containing two copies of the 136 bp Hha 1 - Stu 1 DNA in a head-to-tail orientation. Lanes: 1, 18 μg of P19 nuclear extract; 2, 36 μg of the same P19 nuclear extract preparation.

(C) Extracts were probed with the 158 bp Bam-Eco fragment of the pLSe140 plasmid. Lanes: 1, 12 μg of P19 nuclear extract (NE); 2, 5 μl each of Fraction 7 and Fraction 8 from heparin-agarose column #2, combined (F7/F8); 3, 5 μl of Fraction 9 from heparin-agarose column #2 (F9).

(D) Extracts were probed with the 158 bp Bam-Eco fragment of the pLSe140 plasmid. Lanes: 1, 7 μg of Fraction 17 from the NE-17 fractionation of P19 nuclear extract exposed to retinoic acid for four days, chromatographed over wheat-germ agglutinin and sephacryl-300 columns (F17); 2, 36 μg of the same P19 nuclear extract used in (B) above; 3, 72 μg of a different preparation of P19 nuclear extract.


A
F7° F7 NF N
116
84
58
48.5
36.5
26.6
1 2 3 4
B
NE
18 ug 36 ug
200
180
116
97
84
68
58
48.5
1 2
C
NE F7/F8 F9
200
97
68
1 2 3
D
F17 NE
7 ug 36 ug 72 ug
180
116
84
58
48.5
36.5
1 2 3

extract from a new preparation (lane 3) had a dark signal around 120 kD and a much fainter one around 70 kD. 36 μg of the same nuclear extract preparation used in experiment #2 (Figure 39B) did not give a signal (lane 2). Fraction 17 (lane 1) contained a 70 kD signal.

Discussion

The results of these experiments suggest that F19 cells have at least two proteins, of about 120 and 70 kD, which are able to interact with the proximal region of the murine pgk-1 UAS.

The 120 kD signal was observed in nuclear extract (Figure 39A, lane 4; Figure 39D, lane 3), Fraction 7 of the heparin-agarose columns #3 and #1 (Figure 39A, lanes 1 and 2), and Fraction 9 of the heparin-agarose column #2 (Figure 39C, lane 3). All fractions were from the 0.24 M KCl elution, but the heparin-agarose column #3 fractionation was not as sharp as that of the column #1 fractionation. Fraction 7 from the column #1 fractionation and Fraction 9 from the column #2 fractionation both contained the only observable R1 binding activity in band shift assays (Chapter 3, Section 3.2.3). This suggests that the 120 kD protein interacts with the R1 site in the UAS.

As all the probes used in these analyses only contained the 136 bp of the UAS between Hha 1 and Stu 1, it was reasonable to conclude that the other DNA-binding proteins of approximately 250 kD and 70 kD interacted with R2, the only other observed binding site in this region of the UAS.

The 70 kD signal was barely detectable in nuclear extract (Figure 39A, lane 4;

Figure 39B, lane 2; Figure 39D, lane 3), but quite evident in Fraction 17 (Figure 39D, lane 1). As Fraction 18 of the same fractionation experiment was used to obtain the R2 footprint, it is likely that the 70 kD binding protein detected in these southwesterns is a component of the R2 binding activity.

The inconsistent presence of the 250 kD signal could have been due to inefficient transfer of the very high molecular weight protein from the gel to the nitrocellulose, or simply failure to electrophorese long enough for the protein to enter the gel. However, neither the 250 kD nor the 100 kD signals observed in experiment #2 (Figure 39B) were observed in any other experiment. As the same concentration of the same nuclear extract preparation was used in experiment #4 (Figure 39D, lane2), and no signal was detected using the 158 bp probe, it is concluded that the 302 bp fragment containing two copies of the Hha 1 - Stu 1 restriction fragment was unable to interact with the same factors as, and was in fact binding with different factors than, the 158 bp probe containing a single copy of the Hha 1 - Stu 1 fragment.

The cluster of signals around 40-50 kD in experiment #1 (Figure 39A, lanes 1 and 4) and its absence in the other experiments suggests that the nuclear extract used in the first southwestern experiment was at least partially degraded. It is therefore tempting to conclude from these experiments that the single DNA-binding activity which was hypothesized to interact at R1 of the murine *Pgk-1* promoter (refer to Chapter 5) is an approximately 120 kD protein, and that at least one of the two DNA-binding activities hypothesized to interact at R2 is an approximately 70 kD protein.

APPENDIX II

Screening an Expression Library

Introduction

The classic method of cloning a gene for a DNA-binding protein, for which the DNA consensus binding site is known, and no antibodies are available, uses DNA-affinity chromatography. This involves growing large quantities of cells, making extract, purifying the protein of interest from a DNA-affinity column, electrophoresing the protein on a gel, purifying the protein from the gel, obtaining a partial amino acid sequence of the protein, and either making an oligonucleotide to the sequence and screening a cDNA library with it under DNA:DNA hybridization conditions, or producing a specific antiserum and screening an expression library.

A newer method, first used by Singh et al (1988), involves the screening of a cDNA expression library with the DNA binding site under DNA:protein binding conditions.

The advantage of the classic affinity chromatography method is that, given enough extract, almost any DNA binding protein can be purified. The disadvantages are (1) that vast quantities of cells are required, and (2) the DNA-affinity column may not succeed if the protein:DNA interaction is of low affinity.

The advantages of screening an expression library are that it is less labour intensive and less costly. The disadvantages are, however, potentially many, and depend on the nature of the protein:DNA interaction, and the use of a prokaryotic system for expressing a eukaryotic protein.

With these considerations in mind, an attempt was made to clone at least one gene encoding a pgk-1 UAS binding protein by screening an expression library, according to the protocol of Singh et al (1988).

Materials and Methods

Libraries. Two libraries were screened. The first was made from P19 cDNA, oligo d(T) primed, and size selected, packaged into lambda gt11. The cDNA was made by John Lu, and packaged into phage by Trudy Marshall. The second library used was made from mRNA from P19 cells exposed to RA for seven days, provided by Dr.R.Slack. cDNA from this oligo d(T) primed mRNA was made by Dr.I.Skerjanc and packaged into lambda gt22A. In this system, the cDNA is fused to the bacterial lacZ gene, the eventual translation product being a fusion protein.

Plating Cells. Bacteria from *E.coli* strain Y1090 were plated. A single colony was picked and grown overnight in 100 ml Luria Broth (10 g/l bacto-tryptone, 5 g/l bacto-yeast extract, 10 g/l NaCl) containing 0.2% maltose and 50 μg/ml ampicillin. The culture was transferred to a tube and spun at 5400 rpm for 10 minutes at 4°C in a Beckman JA-17 rotor. The cell pellet was resuspended in 20 ml of 10 mM $MgSO_4$. At an optical density at 600 nm of approximately 2, the cell suspension was returned to a flask at 37°C for one hour. The suspension was then stored at 4°C.

Phage Titration. 990 μl of SM medium (10 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl, pH 7.5, 2% gelatin) was delivered to three microfuge tubes. 10 μl of the primary library stock were put in tube 1, to become the 10^2 stock. 10 μl of the 10^2 stock were transferred to tube 2, to become the 10^4 stock. 10 μl of the 10^4 stock were transferred to tube 3, to become the 10^6 stock. 100 μl of the 10^2 stock were added to 900 μl of SM medium to become the 10^3 stock. 100 μl of the 10^4 stock were added to 900 μl of SM medium to become the 10^5 stock.

3 ml of top agarose (0.7 g/100 ml LB) were placed in a 5 ml polystyrene tube at 47°C in a water bath. 1 μl of 100 mg/ml ampicillin and 30 μl of 1 M $MgSO_4$ were added. 100 μl of the 10^3 stock were added to one tube (a 10^4 dilution). 100 μl of the 10^4 stock were added to another tube (a 10^5 dilution). 100 μl of the 10^5 stock were added to a third tube (a 10^6 dilution). 100 μl of plating cells were added to each tube of virus, vortexed, and placed at 37°C. After 20 min the plating cells and bacteria were mixed with the top agarose and plated onto four-day-old plates containing bottom agar (15

g/1000 ml LB). The plates were placed at 37°C overnight. The 10^4 dilution yielded 300 plaques, the 10^5 dilution yielded 30 plaques and the 10^6 dilution yielded no plaques. 300 x 10^4 = 3 x 10^6 plaque forming units (pfu)/ml. Therefore, 100 μl of original stock contained 3 x 10^5 pfu/100 μl.

Probe. The library was screened with either the 158 bp Bam-Eco fragment of pLSe140 or the 302 bp Bam-Eco fragment of pLSe140x2 (for plasmid details refer to Chapter 2). Plasmid, digested with Bam H1 and Eco R1, and dephosphorylated, was labelled with (γ-^{32}P)ATP using T4 polynucleotide kinase. The labelled DNA fragments for probing were gel purified from the vector and any unincorporated nucleotide triphosphate.

Primary Screen. From the first library of P19 cDNA, 3 x 10^5 plaques were screened. 33,000 pfu were plated per 150 mm petri dish, for a total of nine plates. Accordingly, 11 μl of phage were added to 200 μl of Y1090 plating bacteria. This mixture was placed at 37°C for 40-50 minutes, then added to 7.5 ml of top agarose containing ampicillin and magnesium. Plates were incubated at 42°C for 3.5 hours until plaques were just visible. 132 mm nitrocellulose filters, pre-soaked for 30 min in 10 mM IPTG and dried for one hour were placed on the phage plates. The filter containing plates were then placed at 37°C for 3.5 hours. Filters were gently removed and placed protein side up in a dish containing binding buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT) for 10 min, with gentle agitation. Buffer was changed and the filters agitated for an additional 10 min. Filters were transferred to BLOTTO (5% milk, 50 mM Tris-HCl,

pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT) in individual dishes and agitated at room temperature for one hour. The filters were rinsed twice for 5 min at room temperature in 50 ml of binding buffer. For probing, 25 ml of binding buffer containing probe at a concentration of 2 x 10^6 cpm/ml, and poly(dI-dC) at 10 μg/ml, were placed in a petri dish. Each filter was passed through the dish, and stacked in a bag to which the probe solution was added. Hybridization was carried out at room temperature with agitation for one hour. Four, 8 min washes were carried out on each filter, individually in 50 ml of binding buffer. Filters were exposed damp, with a screen, for 13 hours at -70°C.

Secondary screen. Twelve plugs were picked from the primary screen, with the wide end of a glass pipette, and placed in 1 ml of SM medium containing 20 μl of chloroform. After remaining at room temperature for an hour, the phage were titrated. 10 μl of the primary stock were transferred to 990 μl of SM medium to yield a 10^2 dilution; 10 μl of this stock were added to 990 μl of medium to yield a 10^4 dilution; 100 μl of the 10^2 stock were added to 900 μl of medium to yield a 10^3 stock, and; 100 μl of 10^4 stock were added to 900 μl medium to yield a 10^5 stock. 100 μl of the 10^3, 10^4 and 10^5 stocks were plated, giving a 10^4, 10^5 and 10^6 plating titration. Plating was carried out by adding phage to 3 ml of top agarose containing ampicillin and magnesium, as in the primary screen. After 3.5 hours of growth at 42°C plaques were visible on plates containing the 10^4 dilution. 82 mm IPTG soaked filters were placed on those plates. Subsequent handling of the filters was identical to the primary screen. The same probe solution was used to screen these filters, having been frozen at -20°C for two days.

Tertiary Screen. Four plugs were picked from the secondary screen. 10^5 and 10^6 dilutions were plated. The probe solution was freshly made, using the 302 bp probe at 1.44 x 10^6 cpm/ml and 38.4 ng/ml.

Quaternary Screen. Ten plugs were individually picked into 1 ml of SM, each plug visually containing only one plaque. The virus was plated at 10^5 and 10^6 dilutions, however only the 10^5 dilution was probed. The probe solution used for the tertiary screen was re-used.

Lambda DNA Preparation. Primary infection: a plug from the quaternary screen was picked into 1 ml of medium. 100 μl of this stock were added to 100 μl of bacterial culture and placed at 37°C for 10 min. 2 ml of broth containing 10 mM $MgSO_4$ were added to the tube which was then incubated, with agitation, for six hours at 37°C. To lyse the cells, 40 μl of chloroform were added, the tube vortexed, and the bacterial debris pelleted by centrifugation at 5000 g for 5 min in a Beckman JA-20 rotor.

Secondary infection: 15 μl of the primary lysate were added to 1 ml of bacterial culture, and adsorption allowed to proceed for 10 min at 37°C. The bacteria were then added to 50 ml of broth containing magnesium, and grown overnight at 37°C. To lyse the bacteria, 200 μl of chloroform were added and 1.5 g of NaCl. The mixture was swirled until the NaCl dissolved. Debris was pelleted at 4200 g at room temperature in a JA-20 rotor.

DNA preparation: The supernatant was added to 5 g of PEG 8000 in a 50 ml

tube, and the tube inverted to dissolve the PEG. The DNA was precipitated at 4°C overnight. DNA was pelleted at 4200 g at 4°C for 30 min. After drying the pellet briefly, it was resuspended in 600 μl of SM medium and transferred to a microfuge tube. 2.5 μl of DNAse 1 (10 mg/ml) were added to eliminate the host DNA, and the tube incubated one hour at 37°C. 2.5 μl of diethylpyrocarbonate were added to inactivate the enzyme.The sample was divided between two tubes. 600 μl of 1 M Tris-HCl, pH 8.5, 0.1 M EDTA and 1% SDS were added to each tube, followed by a 5 min incubation at 65°C. DNA was precipitated by the addition of 300 μl of 5 M KOAc, pH 4.8, and incubation on ice for 30 min. The debris was spun out in a microfuge for 30 min at 4°C. Isopropanol was added to the supernatant, which was vortexed and left on ice for 30 min. After centrifugation for 4 min at 4°C, the DNA pellet was resuspended in 200 μl TE. Samples were pooled, extracted once with phenol and chloroform, and ethanol precipitated. Following centrifugation, the DNA pellets were resuspended in 200 μl TE.

Characterization of Clone. Phage DNA was digested with Eco R1 to determine the presence of insert DNA. Insert DNA was sequenced in the phage using the lambda gt11 forward primer. Northern analysis was performed on P19 and HeLa RNA using random primed insert DNA as probe. As a control of the Northern analysis, CTF-1 and PEA-3 DNA probes were also used. Southern analysis was also performed, using P10 genomic DNA probed with labelled insert DNA. Genomic P10 DNA was digested with Bam H1, Eco R1 or Hind III prior to electrophoresis through a 1% agarose gel.

Results

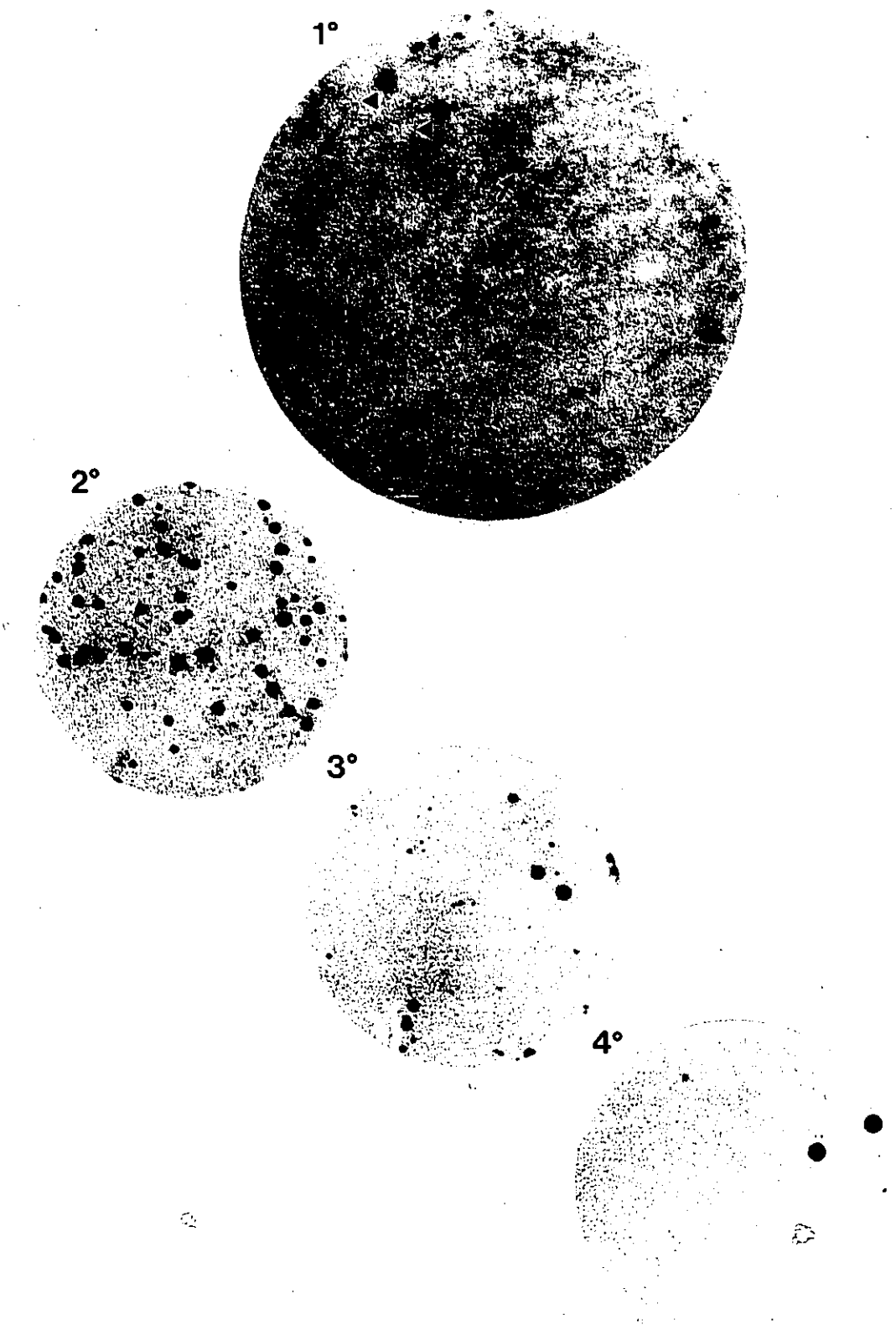
Screening. 3×10^5 plaques of the lambda gt11 library were screened. Working under the assumption that there are approximately 50,000 different transcripts per mouse cell, and roughly ten transcripts of an abundant transcription factor per mouse cell (Dr.Gray, personal communication), of the 3×10^5 plaques, $[(3 \times 10^5)/50{,}000] \times 10 = 60$ would be expected to contain the gene for a particular, abundant, transcription factor. However, considering there is 1/2 a chance of a transcript being in the correct orientation for translation, and 1/3 of a chance of it being in frame, there is a $1/2 \times 1/3 = 1/6$ chance of a particular cDNA being expressed. $1/6 \times 60$ plaques $= 10$, the number of expected positive signals from the screening of 3×10^5 plaques for an abundant transcription factor.

In the primary screen of the P19 cDNA library there were 12 potentially positive signals. (Duplicate lifts were not performed.) 12 plugs, incorporating the potentially positive signals, were used for the secondary screen. Only one of the 12 plates used in the secondary screen had any signal, and on the almost confluent plate there were 61 signals. Four plugs were taken, but the results of the tertiary screen were not as striking, there being many positive signals on every plate, many of which did not correspond to plaques. Ten plugs were removed for the quaternary screen. One plate was positive for every plaque (of which there were only two) (Figure 40).

Roughly 1×10^6 plaques of the P19+RA library were screened. Results varied depending on the wash stringency of the primary screen. There were either no signals,

FIGURE 40: Screening of lambda gt11 P19 cDNA expression library

Three plaques from the 1° screen (indicated by arrows) were re-screened in a 2° screen. Only one placque (arrow with tail) gave any signal in the 2° screen. By the 4° screen there were two placques on the plate, both of which produced a signal.


1°
2°
3°
4°

or every plaque lit up. Secondary screens were performed but with similar results.

Therefore, screening of the P19+RA cDNA library was unsuccessful, but screening of the P19 cDNA library resulted in the identification of two clones expressing protein capable of binding the murine *Pgk-1* proximal UAS sequence.

Eco R1 Digestion. The cDNA for the P19 library was size selected between two and three kb, and Eco R1 linkers were added prior to packaging. It was therefore conceivable that more than one Eco R1 linkered cDNA fragment could insert into the lambda vector.

The digestion of purified phage DNA from both positive plaques from the P19 cDNA library resulted in the release of two Eco R1 inserts, of 2.5 and 2.2 kb. Sequencing of the first 200 bp of recombinant DNA fused to the end of the lacZ gene (using the lambda gt11 forward primer), and comparison to Gene Bank sequences, revealed 84% homology at the amino acid level and 91% homology at the nucleotide level to the gene encoding *E.coli* racC, a protein of unknown function (Chu et al 1989). Sequencing of from 200-400 nucleotides of the other end of the Eco R1 insert fragment, as well as both ends of the other Eco R1 insert fragment did not reveal any more homologies.

Northern Analysis. Northern analysis was performed in order to determine the presence of P19 mRNA which was able to hybridize with insert cDNA. This would indicate that the insert DNA was eukaryotic in origin. Using a 1.5 kb Pst fragment from the 2.5 kb Eco R1 insert and a 1.7 kb Bam H1 fragment from the 2.2 kb insert, no

signal was observed, although the control probes CTF-1 and PEA-3 were able to detect signal in HeLa and P19 RNA samples, respectively. Insert DNA was therefore either prokaryotic or unable to detect a low abundance message in P19's.

Southern Analysis. Southern analysis was performed in order to see if genomic sequence could be detected which hybridized to insert DNA. As in the Northerns, no signals were detected. The cells, therefore, did not appear to contain any gene encoding message from which the insert DNA was made.

Sequence specificity. The purified phage containing the 2.5 and 2.3 kb inserts, and a purified phage which did not produce any signal during the screening procedure, were tested for binding specificity using the 130 bp Hind III - Pvu II AdMLP DNA fragment. Plaques from the 2.5 and 2.3 kb insert containing phage all produced signal, whereas no signals were observed on the plate containing plaques from the control phage. This suggested that the *E.coli* protein encoded by the *racC* gene may be a non-specific DNA-binding protein.

Discussion

The DNA sequence fused to the 3'-end of the lacZ gene had 91% homology to the *E.coli racC* gene (Chu et al 1989). This result, in conjunction with the fact that neither the 2.5 nor 2.3 kb phage inserts hybridized to mouse RNA or DNA in Northern and Southern analyses, suggests that the origin of the insert DNA is contamination of bacterial DNA

during packaging (Huynh et al 1985). The results of the test for sequence specificity indicate that the racC protein was detected during the screening procedure due to its high affinity, non-specific DNA-binding characteristics.

Failure to detect a P19 UAS binding protein in the expression library could be due to a number of reasons. First, perhaps no pgk-1 UAS binding protein binds the DNA with high enough affinity to withstand the washing procedure. To increase the affinity of the interaction, a probe containing multiple repeats of the same consensus site, on concatenated DNA fragments, can be used so that the probe simultaneously binds two or more immobilized protein molecules (Vinson et al 1988; Singh et al 1988). The 302 bp Bam-Eco fragment of the pLSe140x2 plasmid contained two copies of the pgk-1 Hha 1 - Stu 1 fragment linked in a head-to-tail orientation, and was used in an attempt to increase affinity of interaction. If this was the reason why no eukaryotic genes were detected, then obviously two copies were insufficient. Second, not enough clones were screened. Based on the above calculations, 3×10^5 clones should have resulted in approximately 10 positive signals for an abundant transcription factor. Fewer positive signals would be expected for a transcription factor which was not abundant. Also, if the factors are abundant because the proteins are not turned over rapidly rather than because the mRNA transcripts are abundant, then the number of expected positive signals would not be as high, and more plaques would have to be screened to increase the likelihood of detection. Third, the tail end of the library being screened no longer contained a representative cross-section of clones. This would most likely only be the case if the library had been re-amplified after a significant number of screenings had already been

performed. This was, apparently, not the case. Fourth, the pgk-1 UAS binding proteins require subunit associations for DNA binding to occur. This explanation is plausible, based on the results presented in Chapters 3 and 4 and the model presented in Chapter 5. Fifth, the DNA binding proteins require modifications in order to bind, modifications such as phosphorylation or glycosylation which are unable to occur in bacterial systems. This explanation is reasonable, as many eukaryotic transcription factors are modified, and some require modification in order to bind DNA (e.g. RAP1) (Chambers et al 1989). Sixth, the bacterially expressed fusion protein may not be appropriately folded and hence unable to recognize the DNA. This may be overcome by denaturing and renaturing the protein, using guanidine hydrochloride (Vinson et al 1988). The limitation of this technique is that not all proteins may be resilient to the chemical denaturation and renaturation procedure.

It is therefore recommended that any future attempt to clone genes encoding pgk-1 UAS binding proteins using the expression library screening method would benefit from screening more plaques and using a concatenated probe. Use of the classic DNA-affinity chromatography method might also be considered, as it would avoid the potential problems concerning lack of post-translational modification associated with a bacterial system, and lack of subunit interaction associated with using an expression library.

APPENDIX III

Papers

Nucleic Acids Research, Vol. 19, No. 20 5755–5761

The mouse Pgk-1 gene promoter contains an upstream activator sequence

Michael W.McBurney, Leslie C.Sutherland, Chakar N.Adra, Benoit Leclair, Michael A.Rudnicki and Karen Jardine

Departments of Medicine, Biochemistry and Biology, University of Ottawa, 451 Smyth Road, Ottawa, Ontario K1H 8M5, Canada

Received May 23, 1991; Revised and Accepted September 23, 1991

ABSTRACT

The *Pgk-1* gene encodes the housekeeping enzyme, 3-phosphoglycerate kinase, and is ubiquitously expressed. This gene resides on the X chromosome in mammals and is always expressed except where it is silenced along with most other genes on the inactive X chromosome of female somatic cells or male germ cells. The *Pgk-1* promoter is in a region rich in nucleotides G and C. This promoter can efficiently drive high levels of expression of reporter genes such as *E.coli lacZ* and *neo*. We have determined that the 120 bp upstream of the transcription start site functions as a core promoter. Upstream of this is a 320 bp region which enhances transcription from the core promoter in an orientation and position independent fashion. This 320 bp region does not enhance transcription from the core promoter of the SV40 early region. Nuclear proteins bind to this 320 bp fragment although the restricted regions to which binding can be demonstrated with gel mobility shift assays suggests that the activity of the enhancer may be mediated by factors which bind at multiple sites each with low affinity.

INTRODUCTION

Most studies on the regulation of gene transcription have focused on those genes whose expression is restricted to specific tissues or is inducible with chemical or physical stimuli. However, the protein products of many genes are required in all cell types and the promoters driving the transcription of these constitutively expressed genes have been studied much less thoroughly. In general, ubiquitously expressed genes have promoters which are rich in G and C and often lack the TATA and CAAT sequence motifs common to promoters of inducible and tissue specific genes. The promoters of some constitutive genes seem relatively simple; for example, fewer than 100 bp of the DNA sequence upstream of the transcription start sites are sufficient to drive efficient expression of the genes encoding hypoxanthine phosphoribosyl transferase (1), adenine phosphoribosyl transferase(2), and thymidylate synthase (3). Recently, however, the promoters of some constitutively expressed genes such as those encoding proliferating cell nuclear antigen (PCNA) (4) and acetyl CoA carboxylase (5) have been shown to be more complex, consisting of both a core promoter along with an upstream region with transcription enhancing activity.

The *Pgk-1* gene of mammals is X-linked and encodes the enzyme 3-phosphoglycerate kinase. This enzyme is involved in the glycolytic pathway and is present in all somatic cells. The PGK-1 protein may also play a role in DNA replication (6). Although the PGK-1 protein is ubiquitous, the *Pgk-1* gene may be either transcriptionally active or inactive. In the somatic cells of female mammals only the *Pgk-1* allele on the active X chromosome is transcribed, the other *Pgk-1* allele on the inactive X being inert. During spermatogenesis, the X chromosome becomes inactivated and transcription of the *Pgk-1* gene ceases (7).

As a first step towards investigating the nature of the inactivation of the *Pgk-1* gene, we have cloned the mouse *Pgk-1*^{b} (8) and *Pgk-1*^{a} (9) alleles which encode two electrophoretically distinct PGK-1 proteins. The promoter of the murine *Pgk-1* gene resembles those of other constitutive genes being GC rich, having no TATA box, but having a CAAT box motif (8). There is a high degree of sequence identity between the human and murine *Pgk-1* promoters. Both human and mouse promoters are unmethylated when active. When inactivated, the promoters become hypermethylated although the extent of methylation of the inactive human *Pgk-1* promoter is much higher than that of the mouse (10–12).

Transfection studies with the *Pgk-1* gene suggested that it is driven by a strong promoter (9). We set out to investigate this promoter in more detail and report below that the murine *Pgk-1* promoter is indeed very active following DNA mediated transfection into recipient cells and is comprised of a core promoter along with an upstream activator region.

MATERIALS AND METHODS

Construction of recombinant genes

The chimeric genes used in these studies were constructed in pGEM3 or 4 by standard procedures (13). All constructs were verified by restriction mapping or DNA sequencing. The *Pgk-1* and *Pgk-1*^{b} cloned genes and their restriction maps and sequences were as described (8,9). We made use of the coding region of an *E.coli lacZ* gene (14) and the *neo* gene from pMC1*neo* (15). The SV40 promoter and enhancer were derived

from pSV1 (16). The plasmid DNAs were isolated using the procedure of Marko et al (17).

Primer extension

Primer extension was performed by a modification of the method of Gerard (18). The 25 nt primer (5'TCCAAAGTCAGCTTG-TTGGAAAGCCG3') was end labeled with T4 polynucleotide kinase and mixed with 250 μg of P19 cell RNA in 25 μl of 50 mM Tris-HCl (pH 8.3) and 8 mM $MgCl_2$. The mixture was boiled for 1 min then incubated at 65° for 5 min, 37° for 10 min, and 20° for 10 min. Then 25 μl of extension buffer was added (50 mM Tris-HCl, pH 8.3, 8 mM $MgCl_2$, 20 mM DTT, 2 mM each of dATP, dTTP, dCTP, and dGTP, and 60 U of Moloney murine leukemia virus reverse transcriptase (Pharmacia)). This mixture was incubated at 20° for 15 min, 37° for 30 min, and 42° for 30 min. The RNA was then degraded by addition of 150 μl of TE buffer containing 5 μg/ml RNase and incubated at 37° for 30 min. The DNA was then extracted and fractionated on an 8% polyacrylamide gel in the presence of 8 M urea.

DNA mediated cell transfection

P19 embryonal carcinoma cells (19) were cultured as described (20) and transfected by the calcium phosphate procedure of Chen and Okayama (21). Transfection experiments were carried out on different DNA preparations in multiply repeated experiments. The activities of β-galactosidase were measured spectrophotometrically (22) and were normalized to the activity from the construct comprised of the *Pgk-1* promoter (−440 to +80) driving *E.coli lacZ*.

DNA binding experiments

Nuclear extracts of P19 cells were prepared as described (23) and used for gel mobility shift experiments (24,25) using DNA fragments end labeled with ^{32}P. For each assay, 20 μg of nuclear protein was mixed with the radioactive probe in the presence of 3 μg of poly(dI.dC).

RESULTS

Mapping transcription start sites

We have previously sequenced the 838 bp of DNA upstream of the first coding exon of the mouse *Pgk-1* gene and noted that this region was GC rich and had a CAAT box sequence but no TATA box (8,9). The absence of a TATA box is often associated with multiple sites of transcription initiation (3). We used the primer extension technique was to map the site(s) of transcription initiation of the *Pgk-1* gene. A 24 nt oligomer complementary to the N-terminal coding region of the mRNA was used as primer and the reverse transcriptase extension products were separated on a gel in which the same primer was used to create a sequencing ladder from the cloned genomic fragment. Five discrete extension products were evident and indicated transcription initiation sites between 101 and 44 bp upstream of the site of translation initiation (Figure 1). Only one of these transcription initiation sites, at −86 bp, conforms to the consensus sequence for eukaryotic initiation (26) and this was the site of one of the strongest signals in Figure 1. Tamaru et al (27) used S1 nuclease mapping with the same promoter and concluded that the major start sites were concentrated near −86 bp. The most upstream initiation site was 101 bp 5' to the translation initiation site and we labeled this site +1 in subsequent figures.

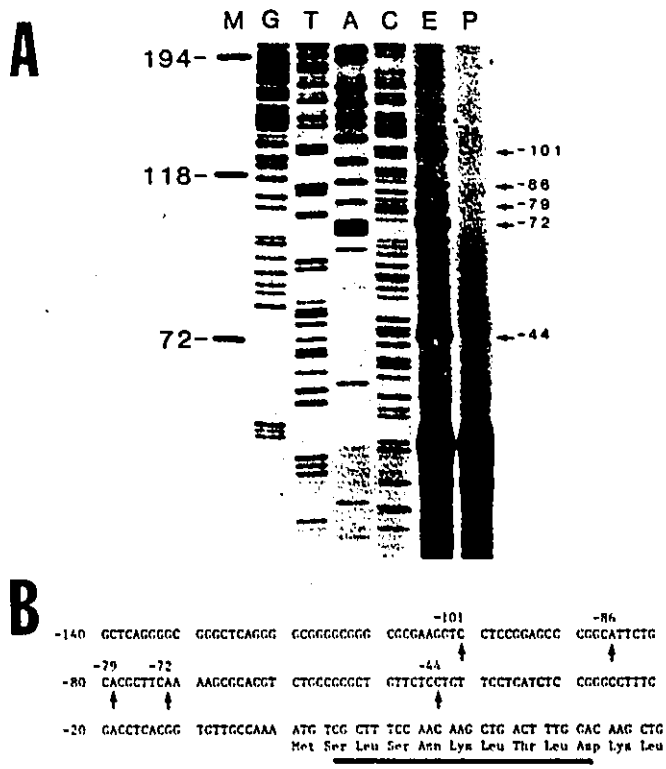


Figure 1. Mapping the sites of transcription initiation of the mouse *Pgk-1* gene. A 25 nt oligodeoxynucleotide complementary to the coding region underlined in panel B was end labeled, hybridized to RNA from P19 cells, and extended with reverse transcriptase. The extension products were separated on a gel and are shown in lane E in panel A along with a sequencing ladder (lanes labeled G, T, A, C) in which the 25 nt oligomer was used to prime sequencing reactions from the cloned promoter region. The reaction product in the absence of RNA template is shown in lane P and DNA size markers are in lane M. The arrows on the right indicate the lengths of the extension products measured in bp upstream of the ATG initiation of translation codon. These are the probable sites of transcription initiation and they are indicated as arrows in the sequence of the 5' region of the *Pgk-1* gene shown in panel B.

Pgk-1 promoter activity

To investigate the activity of the *Pgk-1* promoter, we constructed chimeric genes in which the coding region of the *E.coli lacZ* gene was inserted at the TaqI site (+80) which is downstream of all *Pgk-1* transcription initiation sites and 20 bp upstream of the *Pgk-1* translation initiation site. The *lacZ* gene was followed by the SV40 polyadenylation and transcription termination signals (Figure 2).

Each construct was transfected into P19 embryonal carcinoma cells and the β-galactosidase activities in these cultures were determined after 48 hr. Strong activity was observed from constructs a, e, and f carrying *Pgk-1* promoters of 4.5 kb, 838 bp and 440 bp respectively. Significantly reduced levels of activity were present in constructs carrying 212 bp and 120 bp of upstream sequence (constructs i and k) suggesting that important components of the promoter reside between −120 and −440 bp. Indeed internal deletion of this 320 bp region, constructs c and d, resulted in a 10 fold reduction in promoter activity.

The 320 bp fragment was cut into two components of 228 bp (−212 to −440) and 92 bp (−120 to −212). Each of these components was assayed independently for activity in constructs

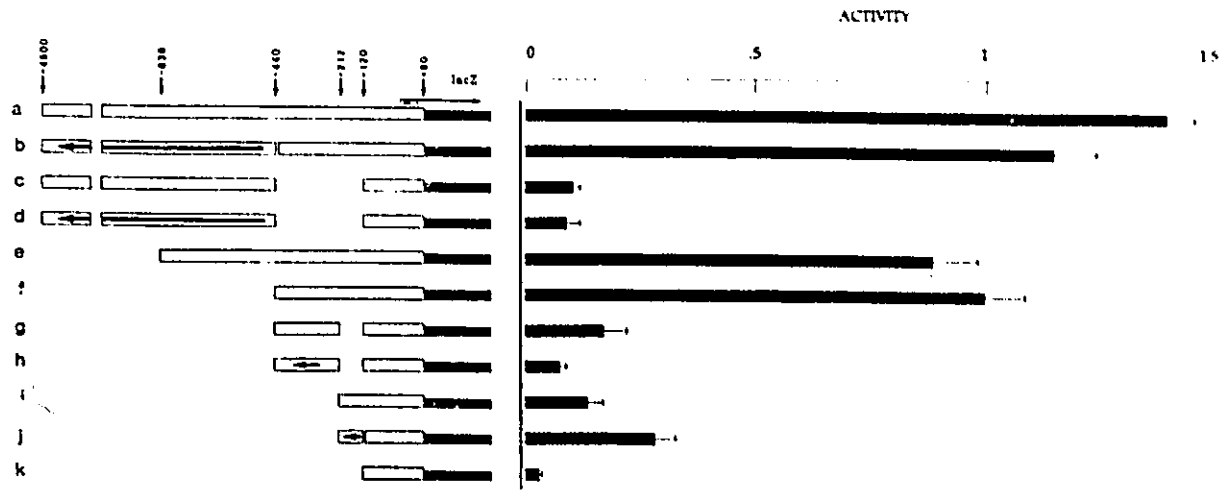


Figure 2. The *Pgk-1* promoter activity in transfection experiments. The coding region from the *E.coli lacZ* gene (black half height box) was fused to the *Pgk-1* genomic sequence downstream of the 5 transcription initiation sites but upstream of the ATG codon. The *Pgk-1* sequences are indicated as open boxes. The *Pgk-1* sequences for constructs e to k (−838 to +80) were derived from the *Pgk-1*^{b} gene (8) while the sequences from −4.5 kbp to −440 bp used in constructs a to d came from the *Pgk-1*^{a} gene (9). Each construct was transfected into P19 cells and 48 hr later the specific activities of β-galactosidase were measured. An average of 9 (between 5 and 14) independent transfection experiments were performed for each construct and at least 2 independent DNA preparations of each construct were tested. The β-galactosidase activities relative to that of construct f (−422 to +80) are indicated to the right of each construct along with bars indicating the standard errors. The arrows in constructs b, d, h, and j indicate that the DNA fragments containing the arrows were inserted in the orientation opposite that of the normal sequence.

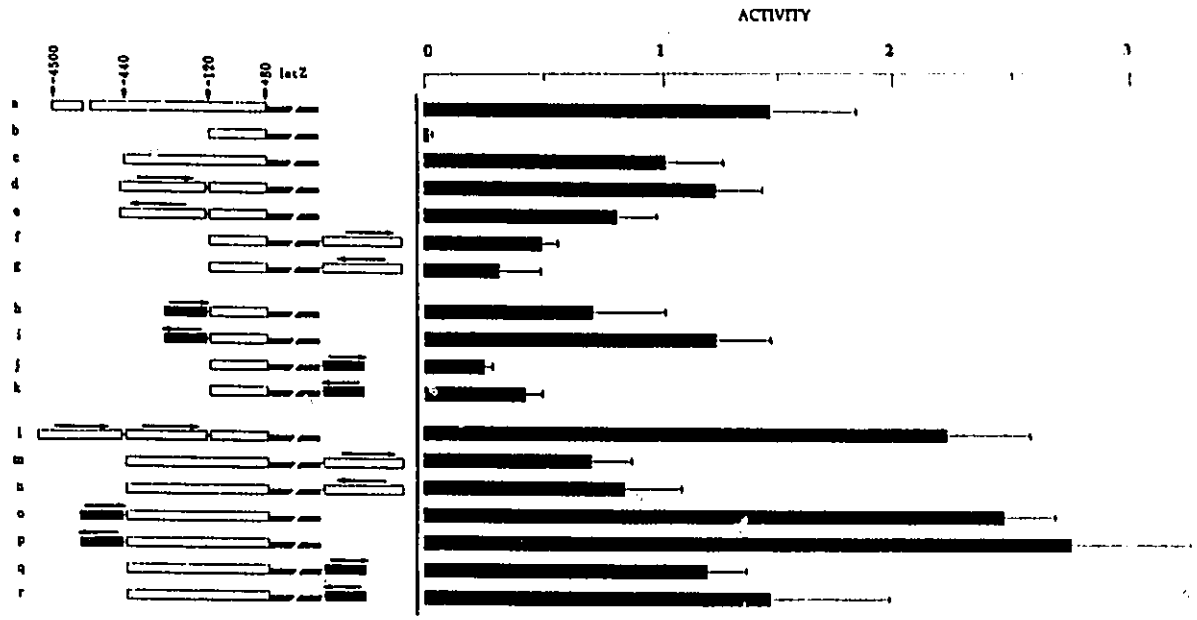


Figure 3. The *Pgk-1* enhancer is orientation independent. All constructs carried the coding region for the *E.coli lacZ* reporter gene followed by the SV40 polyadenylation and transcription termination signals (black half height boxes). The DNA sequences derived from the *Pgk-1* promoter are shown as open boxes. The 320 bp open boxes with arrows above them shown in constructs d, e, f, g, l, m, and n are from the −440 to −120 bp region and the orientation is indicated by the arrow (natural orientation is with arrow pointing right). The 166 bp hatched boxes in constructs h, i, j, k, o, p, q, and r contain the enhancer of the SV40 early promoter in the orientation indicated by the arrows. The thin lines between boxes indicate short (less than 15 bp) linker DNA sequences used in constructing the composite genes. The activities of each construct are shown on the right relative to the activity of construct e (−422 to +80) along with the standard error of each. Each construct was tested in between 3 and 10 independent transfection experiments.

g, h, i, and j. Each of these two components enhanced the activity of the basal promoter (construct k) regardless of orientation but the extent of enhancement of either fragment was slight compared to that achieved by the intact 320 bp fragment (compare the activities of construct f with those of constructs g to j). This result suggests that the full enhancement from the 320 bp fragment

requires cooperation between factors bound to regions on both sides of the AluI restriction site (−212) used to cut this fragment in two or that an important DNA sequence element at or near this AluI site is destroyed by restriction digestion.

To determine if cellular factors bind to and activate the upstream sequences and could be competed by an excess of this upstream DNA (28), construct f was cotransfected with a 10 to 50 fold molar excess of plasmid containing the 320 bp fragment (−440 to −120). Expression of β-galactosidase from construct f was not diminished by the presence of excess 320 bp fragment (data not shown) suggesting that the cellular factors which bind to the 320 bp region and enhance the activity of the *Pgk-1* promoter are either present in abundance in P19 cell nuclei or they bind only in the presence of an adjacent promoter.

The 320 bp fragment enhances regardless of orientation and position

The core promoter used in these studies, −120 to +80, contains four potential Sp1 binding sites and a CAAT box (see Figure 7). The 320 bp sequence upstream of this core promoter was inserted in either orientation both upstream and downstream of the core promoter (Figure 3, constructs d, e, f, and g). Elevated expression of the *lacZ* reporter gene was evident in all four constructs indicating that the 320 bp sequence functions in an orientation independent fashion to enhance promoter activity. The constructs in which the 320 bp fragment was inserted downstream of the reporter gene were less active than those in which this fragment was inserted proximal to the core promoter indicating that the enhancing activity of this 320 bp fragment decreases with distance from the promoter.

In constructs h, i, j, and k a 166 bp fragment containing the enhancer of the SV40 early promoter was inserted both upstream and downstream of the core *Pgk-1* promoter construct shown in b. The effect of the SV40 enhancer was virtually the same as that of the 320 bp *Pgk-1* derived fragment--the activity of the core promoter was elevated regardless of enhancer orientation and the level of expression was higher when the enhancer was proximal to the promoter.

The effect of two enhancing sequences was investigated in constructs l to r. Tandem upstream enhancing fragments seemed to have an additive effect (constructs l, o, and p) regardless of whether the enhancing fragment was of *Pgk-1* or SV40 origin. Addition of a second enhancing sequence downstream had no additional enhancing effect.

Pgk-1 enhancer does not activate the SV40 core promoter

In the constructs shown in Figure 3, the behaviour and strength of the 320 bp sequence from the *Pgk-1* upstream region was similar to that of the SV40 enhancer. To determine whether the 320 bp fragment could activate transcription from the SV40 core promoter the constructs shown in Figure 4 were made. The 320 bp *Pgk-1* fragment had very little effect on the activity of the SV40 core promoter (constructs d, e, and f) compared to the activity of the intact SV40 early promoter carrying its enhancer (construct a). The 320 bp fragment had no effect on the intact enhancer-containing SV40 promoter (construct b).

We investigated the possibility that an enhancer sequence might exist in the *Pgk-1* promoter upstream of the 320 bp fragment but we found no evidence for the presence of an enhancer in this 4.1 kb region (constructs g to j).

The constructs shown in k and l show the activities of the *Pgk-1* promoter of 440 bp and 120 bp respectively. The intact *Pgk-1* promoter (construct k) has activity comparable to that of the intact SV40 early promoter (Figure 3). Although the activity of the

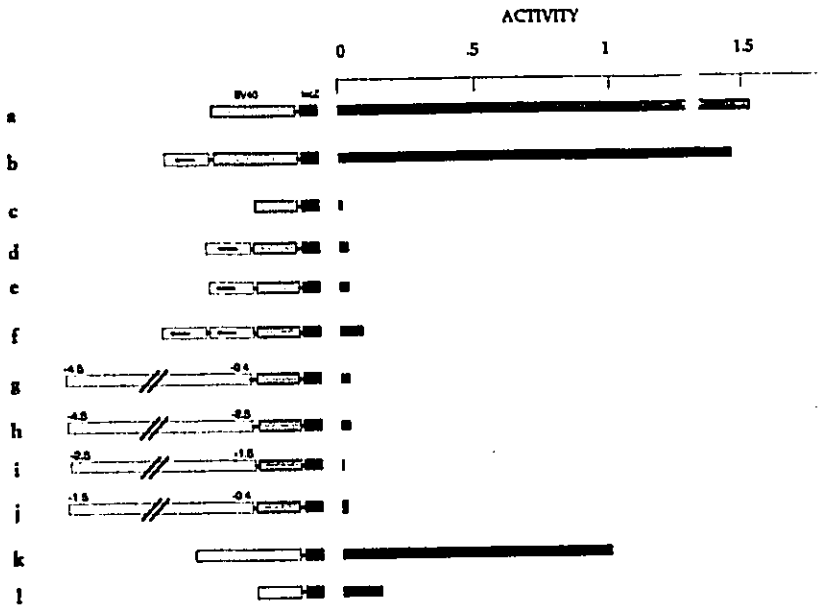


Figure 4. The *Pgk-1* enhancer does not activate expression from the SV40 core promoter. The *E.coli lacZ* reporter gene in the black box was driven by the 350 bp SV40 early promoter (a) with its enhancer region intact or by the 135 bp core promoter (c) lacking the enhancer. The SV40 sequences are hatched and the *Pgk-1* derived sequences are open boxes. Those open boxes in constructs b, d, e, and f are the 320 bp *Pgk-1* enhancer oriented in the direction indicated by the arrows. In constructs g, h, i, and j the *Pgk-1* derived sequences are from the indicated regions in kbp upstream of the enhancer. Constructs k and l are the *Pgk-1* promoter of 440 and 120 bp respectively. The activities of β-galactosidase are averages from between 3 and 6 independent experiments and are represented as the shaded bars to the right of each construct relative to construct K (−422 to +80).

Pgk-1 core promoter was enhanced by both the 320 bp *Pgk-1* fragment and SV40 enhancer, the activity of the SV40 core promoter was not elevated by the 320 bp *Pgk-1* fragment. This is somewhat surprising because both core promoters are GC rich and carry multiple Sp1 binding sequences.

Protein binding to the 320 bp fragment

To determine whether nuclear proteins are capable of binding to the 320 bp *Pgk-1* fragment, we extracted proteins from the nuclei of P19 cells and used these in gel mobility shift assays (24,25). The 320 bp fragment yielded two retarded bands (Figure 5, lane 1). The binding activities of various subfragments of the 320 bp region were assessed as shown in Figure 5. A number of fragments were capable of yielding these two retarded complexes. The nature of the proteins responsible for these two bands has not yet been established. The smallest fragment which gave rise to the 2 retarded bands was a 30 bp sequence shown in panel 11. This sequence is indicated in Figure 7 and contains an E-box (CANNTG), the site to which the helix loop helix class of proteins bind (29). This 30 bp binding site lies within the fragment, −120 to −212 bp, shown to have some promoter enhancing activity (Figure 2 constructs i and j). Little or no binding activity was detected with DNA probes spanning the AluI site at −212 (Figure 5, lanes 5 and 10) or with probes upstream of this site (lanes 2 and 4). Because the activity of the 320 bp fragment could not be mimicked with shorter fragments it seems likely that there are proteins in the nuclei of P19 cells which interact with sequences upstream of the 30 bp region but which are not detected in the mobility shift assays, perhaps because their binding affinities are low or because they are not efficiently extracted from nuclei with the procedures we used.

The two bands detected in the mobility shift assays result from protein(s) which appear to bind specifically because unrelated sequences were incapable of competing for binding activity while the binding complexes were efficiently competed by an excess of any of the fragments labeled as positive in Figure 5 (data not shown).

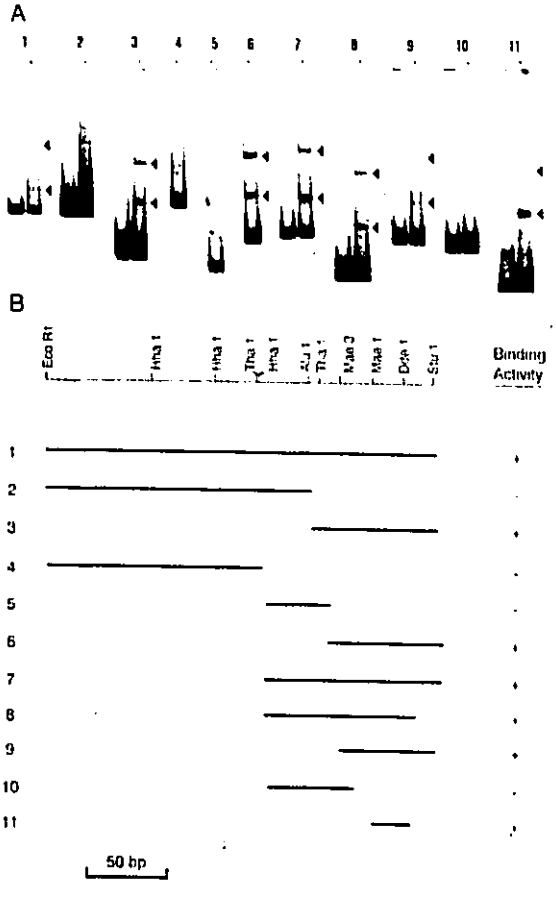


Figure 5. Nuclear proteins bind the *Pgk-1* enhancer DNA. The DNA fragments used for the gel shift experiments are indicated in panel B. The EcoRI (−422) and StuI (−120) sites span the *Pgk-1* enhancer. The presence of detectable binding activity is indicated to the right of each fragment. In panel A the autoradiograms of gel shift experiments is shown for each of the 11 DNA fragments. These lanes have been exposed to x-ray film for different periods of time. In each case + and − indicate the presence and absence of nuclear extract respectively. Arrowheads indicate the two retarded complexes found using those fragments with strong DNA binding activity. Retarded bands were occasionally seen with fragments labeled negative after long exposures (fragments 2, 4, 5, and 10) but these appeared to be due to effects or factors distinct from those which give rise to the pair of complexes on fragments labeled postive.

Selectable gene constructs

Because of the ubiquitous expression of the *Pgk-1* gene and the apparent strength of its promoter, we examined the possibility that the *Pgk-1* promoter might be useful in strong universal expression vectors. We constructed chimeric genes comprised of the *neo* gene derived from pMC1*neo* (15) inserted behind the −440 to +80 *Pgk-1* promoter and upstream of sequences derived from the 3′ end of the *Pgk-1* gene. These 3′ sequences included the polyadenylation signal (AATAAA), the site of polyadenylation, and presumptive transcription termination sequences (9). These constructs (Figure 6) carrying 1.5 kbp (construct A), 600 bp (construct B), and 292 bp (construct C) of 3′ sequence were transfected into P19 cells along with

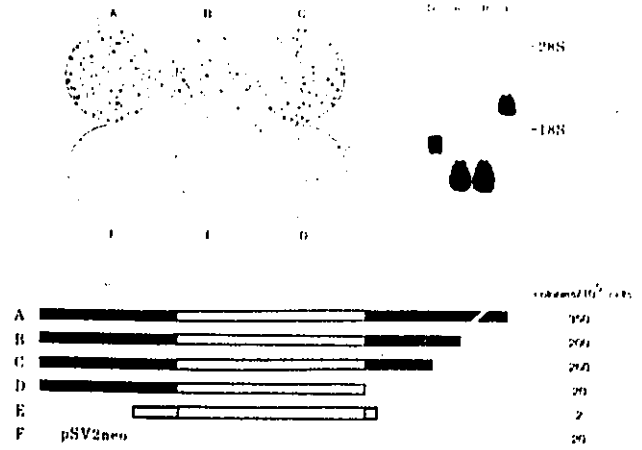


Figure 6. The *Pgk-1* 5′ and 3′ flanking regions drive efficient expression of the *neo* gene. In the lower panel, constructs A, B, C, and D consist of the −422 to +80 bp *Pgk-1* promoter (black box) driving the *neo* gene (open box) derived from pMC1*neo*pA (15) (construct E). The 3′ sequences from *Pgk-1* start at the PvuII site present 28 nt upstream of the polyadenylation site and extending for 1.5 kbp (A), 600 bp (B), 292 bp (C), or 0 bp (D). Construct F is pSV*neo* (52). The number of colonies formed in G418 (400 μg/ml) is the average of 2 experiments. The upper left panel shows representative plates stained 7 days after 100,000 cells were seeded following transfection with each of the 6 constructs. The Northern blot (upper right) shows the RNA from pooled populations of G418 resistant colonies probed for the *neo* sequence.

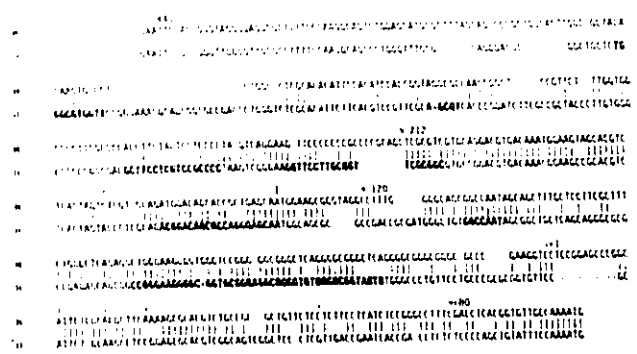


Figure 7. Comparison of the mouse (M) and human (H) *Pgk-1* promoter sequences. The sequences span the mouse sequence −422 to +101, the site of the ATG codon. Sites of nt identity are indicated by perpendicular lines. The arrows on the mouse sequence indicate the sites of transcription initiation and the numbered triangles indicate sites used in this study. The 30 bp region spanning nt −165 to −135 is the shortest DNA fragment used in Figure 5 for gel shift experiments. On the human sequence, highlights indicate the sites of DMS (dimethyl sulfate) footprints seen *in vivo* by Pfeifer et al (38).

pMC1*neo*pA and pSV2*neo*. Colonies were selected for growth in G418. The *Pgk-1* driven constructs A, B and C gave rise to more than ten times as many colonies as either pMC1*neo*pA or pSV2*neo* and the colonies formed following transfection of constructs A, B, and C grew more rapidly. RNA isolated from pooled populations of G418 resistant clones carried abundant *neo* mRNA. These *Pgk-1*-driven *neo* constructs were efficient only in the presence of downstream polyadenylation and transcription termination signals (construct D).

The *neo* gene from pMC1*neo* was used for the *Pgk-1* driven constructs A to D and it carries a mutation which results in reduced phosphotransferase activity (30). When replaced with the wild type *neo* gene present in pSV2*neo*, construct B was an additional 3 to 5 times more efficient in giving rise to G418 resistant colonies (data not shown).

The *Pgk-1*-driven *neo* constructs shown in Figure 6 have been found to be efficient in a variety of cell lines including NIH3T3, rat L6 myoblasts, V79 Chinese hamster fibroblasts, and a variety of mouse embryonal carcinoma and embryonic stem cell lines. Highly efficient expression of various other genes or cDNAs has been induced by the *Pgk-1* promoter and 3′ sequence used for constructs B and C; these include genes encoding hygromycin resistance, *E.coli lacZ*, retinoic acid receptors cDNAs (31) and *Pgk-1* pseudogenes (9).

DISCUSSION

The PGK-1 enzyme is essential in all cells because it is necessary for glycolysis. The *Pgk-1* gene is therefore expressed at high levels in virtually all cell types. In yeast the promoter of the *Pgk-1* gene is one of the strongest known in that species (32). The activity of the mouse *Pgk-1* promoter including its enhancer element is comparable to those strong viral promoters such as SV40 and RSV. Because it is active in virtually all somatic and germ cell types, the *Pgk-1* promoter should be useful for driving other genes at high constitutive levels. In this regard it has been useful in driving sense and antisense cDNA constructs (33), in retroviral constructs (34), and in constructs for selecting homologous recombinants in embryonic stem cells (Rudnicki, unpublished).

In transient transfection assays the *Pgk-1* and SV40 early promoters appeared to drive transcription at comparable rates; however, the SV40 driven *neo* gene was very much less effective than the *Pgk-1* driven *neo* in giving rise to G418 resistant colonies of P19 cells. The reason for this discrepancy between transient and stable expression is not yet clear. We have noticed that cells transfected with *neo* genes segregate G418 sensitive cells when grown in the absence of the drug. One possible explanation for the difference in both colony number and size is that the SV40 promoter becomes inactivated following integration into the genome at a rate much higher than the *Pgk-1* promoter.

Our transfection assays with constructs driven by the *Pgk-1* promoter have indicated the presence of a 320 bp sequence upstream of the core promoter which enhances transcription from the core promoter in an orientation and position independent fashion. Both the core promoter and enhancer element of the *Pgk-1* upstream sequence are part of a GC rich island (35). The DNA sequence upstream of the human and mouse *Pgk-1* genes is highly homologous and is shown and compared in Figure 7. Our work using gel shift assays has indicated that nuclear proteins can bind to sites near the 3′ end of the 320 bp enhancer. The 30 bp fragment found effective for mobility shift assays is indicated on Figure 7.

The activities of well-characterized promoters and enhancers are dependent on binding multiple transcription factors (36,37), yet we have detected binding of factors only to the 3′ end of the 320 bp enhancing region. Pfeifer et al., (38) have investigated the human *Pgk-1* promoter by *in vivo* footprinting with DMS and more recently with DNaseI (39). Their work has indicated five footprints in the region corresponding to the 320 bp enhancer (−120 to −420) and these are indicated in Figure 7 by highlights on the human DNA sequence. There is overlap between one of the human footprints and the 30 bp fragment used in our mobility shift assays but none of the proteins which gave rise to the other *in vivo* footprints on the human *Pgk-1* promoter were able to form DNA protein complexes which we could detect *in vitro* on the homologous mouse *Pgk-1* sequence. There are a number of possible explanations for our failure to detect DNA binding to these 5′ regions of the 320 bp fragment. We favour the idea that binding to these sites involves protein complexes which have relatively low affinity for the DNA and that *in vivo* there is cooperation between the binding factors enabling the establishment of a stable active DNA configuration (40). Cooperative binding of transcription factors is consistent with the observation (Figure 2) that the activity of the 320 bp *Pgk-1* enhancer is substantially reduced in constructs in which either the 5′ or the 3′ part of this 320 bp region is deleted.

Although the *Pgk-1* gene is expressed in all somatic cells the *Pgk-1* allele on the inactive X chromosome in female somatic cells is transcriptionally silent. Pfeifer et al., (38) found that the human *Pgk-1* gene from the inactive X chromosome carries no footprint on its promoter. We have female embryonal carcinoma cell lines containing two X chromosomes. One of these cell lines contains 2 active X chromosomes (41) while in another 1 of the 2 X chromosomes is inactive (42). Nuclear extracts from these various cells behave the same as those of the P19 cell (which has only one X chromosome) and give rise to similar gel shifts (data not shown). In addition, we have transfected the various *Pgk-1* driven *lacZ* constructs into these female embryonal carcinoma cells and found that the 320 bp fragment functions as an enhancer in all cells tested.

A factor which specifically binds methylated GC islands might

play a role inactivating genes such as *Pgk-1* on the inactive X chromosome; however, this binding activity appears to be low in embryonal carcinoma cells (43) and there is inactivation of X-linked genes in the absence of detectable DNA methylation (44). We favour models of X inactivation in which there is competition between activators and repressor for regulatory sites on each X-linked gene (40). There are precedents for competition between activators and repressor proteins in other genes (45–49). It is perhaps pertinent that the 30 bp sequence used for mobility shift assays contains an E-box (CANNTG), the DNA sequence to which DNA binding proteins of the helix-loop-helix variety bind. Since dosage compensation of X-linked genes in *Drosophila* apparently involves the interaction of helix-loop-helix proteins (50,51), the presence of the E-box in the *Pgk-1* enhancer may not be fortuitous.

ACKNOWLEDGEMENTS

We thank Drs. J.Kralova, W.Colledge, P.Boer and R.Hawley for help of various kinds in creating the constructs used in the experiments reported above. This work was supported by a grant from the Medical Research Council of Canada. MWMcB is a National Cancer Institute of Canada Terry Fox Cancer Research Scientist and Drs. Adra and Rudnicki were supported by studentships from the National Cancer Institute of Canada and the Medical Research Council of Canada.

REFERENCES

1. Melton, D. W., Konecki, D. S., Brennand, J. and Caskey, C. T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2147–2151.
2. Park, J-H. and Taylor, M. W. (1988) *Mol. Cell. Biol.*, **8**, 2536–2544.
3. Deng, T., Li, Y., Jolliff, K. and Johnson, L. F. (1989) *Mol. Cell. Biol.*, **9**, 4079–4082.
4. Pietrzkowski, Z., Alder, H., Chang, C.-D., Ku, D.-H. and Baserga, R. (1991) *Exp. Cell Res.*, **193**, 283–290.
5. Luo, X. and Kim, K. H. (1990) *Nucleic Acids Res.*, **18**, 3249–3254.
6. Jindal, H. K. and Vishwanatha, J. K. (1990) *J. Biol. Chem.*, **265**, 6540–6543.
7. Gartler, S. M. and Riggs, A. D. (1983) *Ann. Rev. Genet.*, **17**, 155–190.
8. Adra, C. N., Boer, P. H. and McBurney, M. W. (1987) *Gene*, **60**, 65–74.
9. Boer, P. H., Potten, H., Adra, C. N., Jardine, K., Mullhofer, G. and McBurney, M. W. (1990) *Biochem. Genet.*, **28**, 299–308.
10. Keith, D. H., Singer-Sam, J. and Riggs, A. D. (1986) *Mol. Cell. Biol.*, **6**, 4122–4125.
11. Bartlett, M. H., Adra, C. N., Park, J., Chapman, V. M. and McBurney, M. W. (1990) *Somatic Cell Mol. Genet.*, **17**, 35–47.
12. Singer-Sam, J., Grant, M., LeBon, J. M., Okuyama, K., Chapman, V., Monk, M. and Riggs, A. D. (1990) *Mol. Cell. Biol.*, **10**, 4987–4989.
13. Maniatis, T., Fritsch, T. and Sambrook, J. *Molecular cloning: a laboratory manual*, Cold Spring Harbor, New York:Cold Spring Harbor Laboratory, 1982.
14. Kothary, R., Clapoff, S., Darling, S., Perry, M. D., Moran, L. A. and Rossant, J. (1989) *Development*, **105**, 707–714.
15. Thomas, K. R. and Capecchi, M. R. (1987) *Cell*, **51**, 503–512.
16. Benoist, C. and Chambon, P. (1981) *Nature*, **290**, 304–310.
17. Marko, M. A., Chipperfield, R. and Birnboim, H. C. (1982) *Anal. Biochem.*, **121**, 382–387.
18. Gerard, G. F. (1987) *Focus*, **9**, 5–6.
19. McBurney, M. W. and Rogers, B. J. (1982) *Dev. Biol.*, **89**, 503–508.
20. Rudnicki, M. A. and McBurney, M. W. Cell culture methods and induction of differentiation of embryonal carcinoma cell lines. In: *Teratocarcinomas and embryonic stem cells, a practical approach*, edited by Robertson, E. J. Oxford: IRL Press, 1987, p. 19–49.
21. Chen, C. and Okayama, H. (1987) *Mol. Cell. Biol.*, **7**, 2745–2752.
22. Norton, P. A. and Coffin, J. M. (1985) *Mol. Cell. Biol.*, **5**, 281–290.
23. Dingnam, J. D., Lebovitz, R. M. and Roeder, R. G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.
24. Fried, M. and Crothers, D. M. (1981) *Nucleic Acids Res.*, **9**, 6505–6525.
25. Chodosh, L. A., Carthew, R. W. and Sharp, P. A. (1986) *Mol. Cell. Biol.*, **6**, 4723–4733.
26. Cereghini, S., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) *Nature*, **209**, 1406–1414.
27. Tamura, M., Nagao, Y., Tana, M., Tadokoro, M., Masamune, Y. and Nakanishi, Y. (1990) *Biochim. Biophys. Acta.*, **1049**, 331–338.
28. Segnan, C., Pellicer, B. K., Carter, A. D. and Hamer, D. H. (1984) *Nature*, **312**, 781–785.
29. Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub, H. and Baltimore, D. (1989) *Cell*, **58**, 537–544.
30. Yenofsky, R. L., Fine, M. and Pellow, J. W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3435–3439.
31. Pratt, M. A. C., Kralova, J. and McBurney, M. W. (1990) *Mol. Cell. Biol.*, **10**, 6445–6453.
32. Odgen, J. E., Stanway, C., Kim, S., Mellor, J., Kingsman, A. J. and Kingsman, S. M. (1986) *Mol. Cell. Biol.*, **6**, 4335–4343.
33. Dinsmore, J. H. and Solomon, F. (1991) *Cell*, **64**, 817–826.
34. Hawley, T. S., Sabourin, L. A. and Hawley, R. G. (1989) *Plasmid*, **22**, 120–131.
35. Bird, A. P. (1986) *Nature*, **321**, 209–213.
36. Lewin, B. (1990) *Cell*, **61**, 1161–1164.
37. Plasne, M. and Gann, A. A. F. (1990) *Nature*, **346**, 329–331.
38. Pfeifer, G. P., Tanguay, R. L., Steigerwald, S. D. and Riggs, A. D. (1990) *Genes Dev.*, **4**, 1277–1287.
39. Pfeifer, G. P. and Riggs, A. D. (1991) *Genes Dev.*, **5**, 1102–1113.
40. McBurney, M. W. (1988) *BioEssays*, **9**, 85–88.
41. McBurney, M. W. and Strutt, B. J. (1980) *Cell*, **21**, 387–364.
42. McBurney, M. W. and Adamson, E. D. (1976) *Cell*, **9**, 57–90.
43. Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. and Bird, A. P. (1989) *Cell*, **58**, 499–507.
44. Bartlett, M. H., Adra, C. N., Park, J., Chapman, V. M. and McBurney, M. W. (1991) *Somatic Cell Mol. Genet.*, **17**, 35–47.
45. Benezra, R., Davis, R. L., Lassar, A., Tapscott, S., Thayer, M., Lockshon, D. and Weintraub, H. (1990) *Ann. NY Acad. Sci.*, **599**, 1–11.
46. Jonat, C., Rahmsdorf, H. J., Park, K. K., Cato, A. C. B., Gebel, S., Ponta, H. and Herrlich, P. (1990) *Cell*, **62**, 1189–1204.
47. Yang-Yen, H.-F., Chambard, J.-C., Sun, Y.-L., Smeal, T., Schmidt, T. J., Drouin, J. and Karin, M. (1990) *Cell*, **62**, 1205–1215.
48. Schule, R., Rangarajan, P., Kliewer, S., Ransone, L. J., Bolado, J., Yang, N., Verma, I. M. and Evans, R. M. (1990) *Cell*, **62**, 1217–1226.
49. Owen, T. A., Bortell, R., Yocum, S. A., Smock, S. L., Zhang, M., Abate, C., Shalhoub, V., Aronin, N., Wright, K. L., Van Wijnen, A. J., Stein, J. L., Curran, T., Lian, J. B. and Stein, G. S. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 9990–9994.
50. Torres, M. and Sánchez, L. (1989) *EMBO J.*, **8**, 3079–3086.
51. Parkhurst, S. M., Bopp, D. and Ish-Horowicz, D. (1990) *Cell*, **63**, 1179–1191.
52. Southern, P. J. and Berg, P. (1982) *J. Mol. Appl. Genet.*, **1**, 327–341.

Intragenic regions of the murine *Pgk-1* locus enhance integration of transfected DNAs into genomes of embryonal carcinoma cells

Michael W. McBurney*, Susan Fournier, Karen Jardine and Leslie Sutherland.

University of Ottawa, Department of Medicine, 451 Smyth Road, Ottawa, Ontario, Canada K1H 8M5

10 Figures
0 Tables

Running title: *Pgk-1* gene transfection

* Corresponding author:

Dr. Michael W. McBurney
Department of Medicine
University of Ottawa
451 Smyth Road
Ottawa, Ontario CANADA
K1H 8M5
Phone: 613-787-6565
Fax: 613-787-6779

SUMMARY

Introduction of recombinant genes into mammalian cells in culture has been an important procedure in establishing the molecular mechanisms of various cellular processes. The efficiency with which recombinant plasmids are expressed following stable integration into genomes of embryonal carcinoma cells is low. Using the P19 embryonal carcinoma cell line as receipients, we found that constructs carrying the promoter and intragenic regions of the murine *Pgk-1* gene were expressed with increased efficiency. This elevated expression was associated with increased numbers of copies of the transfected plasmid DNA stably integrated into the genomes of recipient cells. The elevated plasmid copy numbers may result from enhanced ligation of transfected plasmids because co-transfected plasmids were also integrated in increased numbers. The enhanced integration and expression of transfected plasmids required active transcription through an intragenic region of *Pgk-1* perhaps resulting in more recombinogenic plasmid DNAs.

INTRODUCTION

The murine *Pgk-1* gene encodes phosphoglycerate kinase, a key enzyme required for glycolysis. *Pgk-1* is expressed in all cells except sperm. The *Pgk-1* promoter has been cloned (1) and is able to actively drive expression of foreign cDNAs and genes (8,33). In earlier studies we investigated the utility of the *Pgk-1* promoter to drive the *E.coli lacZ* reporter gene in transgenic mice but found undetectable expression in most lines and most tissues (T. Wilkie, C.N. Adra and M.W. McBurney, unpublished results). More recently we established transgenic strains carrying a construct (*Pgk-1,2-lacZ)* comprised of the *Pgk-1* promoter and its first two introns upstream of the *lacZ* reporter gene. Two of three transgenic strains expressed high levels of *lacZ* (McBurney et al., submitted). *In situ* staining for β-galactosidase expression indicated that there was unexpected cell-to-cell variation in the expression of the *Pgk-1,2-lacZ* transgene suggesting that there may be intragenic regions of the *Pgk-1* gene that play roles in regulating its activity.

Intragenic regulators of gene expression have been identified in a number of genes such as those encoding adenosine deaminase (3,4), proliferating cell nuclear antigen (43) growth hormone (50), thymidylate synthase (5), histone H3 (17), troponin I (23) and *hprt* (48). Like *Pgk-1*, *hprt* is an X-linked gene and efficient expression of its promoter in embryonic stem cells occurs only if constructs contain sequences derived from the first two introns

(48). We found that the activity of the *Pgk-1* promoter was not effected by intragenic regions in transient transfection assays in embryonal carcinoma cells but that the intragenic regions enhanced by 30 fold the expression of the same genes following their stable integration into the genomes of embryonal carcinoma cells.

MATERIALS AND METHODS

Cell culture and transfection

P19 embryonal carcinoma cells (31) and V79 Chinese hamster cells were cultured as described (49).

DNA mediated transfection was performed using the BES-buffered calcium phosphate procedure (9) with the following modifications. The BES-buffer was adjusted to pH 6.88 and 1.0ml of transfection solution containing a total of 12μg of DNA was added to cultures containing 10^6 cells in 5ml of medium in 60mm diameter dishes. In experiments reported in Figures 1-5, the DNA used for transfections consisted of a mixture of 10μg of the test plasmid and 2μg of plasmid carrying a selectible gene, neomycin (54), hygromycin (28), or puromycin (58). In each case the selectible gene was inserted behind the *Pgk-1* promoter and in front of the *Pgk-1* polyadenylation signal (33). In experiments reported in Figures 6-10 the transfected DNA consisted of 4μg of plasmid bearing *Pgk-lacZ*, 4μg of *Pgk-neo*, and 4μg of plasmid carrying B17 or a fragment of B17 or the plasmid vector alone. When comparisons of P19 and V79 cells were made, aliquots of the same DNA calcium phosphate co-

precipitate were added to the cultures of the two cell lines.

The DNA-calcium phosphate co-precipitate was incubated with the cells for 7-8hrs before being replaced with fresh medium. Incubation was continued for 40hrs after which the cells were harvested by trypsinization. An aliquot 10^6 cells from each culture was plated into a 150mm diameter culture dish in 50ml of medium supplemented with antibiotic G418 (400μg/ml), hygromycin (200μg/ml), or puromycin (2μg/ml). The remaining cells were pelleted, frozen at -20°C and later extracted for β-galactosidase determination.

Cultures plated into medium containing G418, hygromycin, or puromycin were re-fed 3 or 4 days following plating. Approximately 1-2 x 10^3 colonies formed in these plates and they were harvested 8 days after plating for β-galactosidase determination. In some experiments these populations of cells were pooled and expanded for 2-4 days for DNA or RNA isolation.

β-galactosidase assays were carried out using a spectrophotometric method (41) with continuous optical density monitoring over 30 min. In each experiment the β-galactosidase activities from different constructs were normalized to that obtained with the *Pgk-1,2-lacZ* construct.

Blot hybridizations

Cellular DNA and RNA were isolated as previously described (6,29). Southern blots of restriction enzyme digested DNA were done as described (29) and Northern blots of RNA were carried out using formaldehyde containing gels (29). DNA slot blots were

performed as described (29). The *E.coli lacZ* and neomycin genes were isolated as restriction fragments from agarose gels and radiolabelled with ^{32}P-dCTP using the multi-prime kit (Amersham). Hybridization, washing and autoradiography were as described (29). Densitometric traces of slot blot autoradiograms were performed with a laser densitometer (LKB).

Recombinant genes

Creation of recombinant genes was accomplished using standard procedures. The B17 fragment of the *Pgk-1* gene is a 17 kbp BamH1 fragment cloned from the *Pgk-1*^{b} gene (2). The *Pgk-1,2-lacZ* construct was created by fusing the *E.coli lacZ* gene to the Nae1 site in exon III. The reading frame was retained with the use of BamH1 linkers. The *Pgk-lacZ* gene consisted of the *Pgk-1* promoter from -440bp (EcoR1) to +80bp (Taq1) relative to the first transcription start site driving the *lacZ* gene terminated by the *Pgk-1* polyadenylation and putative transcription termination signals contained in a 460bp Pst1-HindIII fragment (33).

Colonies were prepared for *in situ* staining of β-galactosidase by fixation for 10 min in phosphate buffered saline containing 0.2% glutaraldehyde then incubated overnight at 37° in X-gal (5-bromo-4-chloro-3-indolyl-β-d-galactoside) (46).

RESULTS

Expression of Pgk-1 constructs containing introns

To investigate the effects of intragenic sequences on the expression from the *Pgk-1* promoter, we constructed a recombinant

gene consisting of the E. coli *lacZ* reporter sequence cloned in frame into the *Pgk-1* third exon. This construct, *Pgk-1,2-lacZ*, is shown in Figure 1 construct *b*. The *Pgk-1* derived sequences comprise those from -860bp to +8kbp relative to the first transcription start site (33) and include the promoter region (-440bp to +80bp) along with introns 1 and 2. The first exon of *Pgk-1* has the translation initiation codons (1) so the properly spliced *Pgk-1,2-lacZ* construct encodes a PGK-β-GAL fusion protein with the first 64 amino acids derived from *Pgk-1* (37) followed by the entire *lacZ* coding region. We compared the expression of the *Pgk-1,2-lacZ* construct with a similar construct lacking intragenic sequences. This construct, *Pgk-lacZ*, consists of the *Pgk-1* promoter (-440bp to +80bp) driving the *lacZ* reporter gene and is shown in Figure 1, construct *a*. The 3′ end of both constructs consist of a 460bp fragment derived from the last (11th) exon of *Pgk-1* and contains the polyadenylation site and putative site of transcription termination (33).

Following transient transfection into P19 embryonal carcinoma cells both *Pgk-lacZ* and *Pgk-1,2-lacZ* yielded similar levels of β-galactosidase activity (open bars in constructs *a* and *b* in Figure 1). In these experiments plasmids carrying the *lacZ* gene were co-transfected with plasmids carrying a gene encoding a selectible marker and cells carrying this marker stably integrated into their genomes were selected in medium containing the appropriate drug. In pooled populations of these stable transformants, cells transfected with *Pgk-1,2-lacZ* contained 20-40 fold more β-

galactosidase activity than cells transfected with *Pgk-lacZ* (solid bars of constructs *a* and *b* in Figure 1). RNA was isolated from these stably transformed cells and Northern blots probed with the *lacZ* sequence. The levels of β-galactosidase expression were proportional to the abundance of the *lacZ* transcripts. A single transcript from the *Pgk-1,2-lacZ* gene was present that was slightly longer than that from the *Pgk-lacZ* gene indicating efficient splicing of the 2 introns from *Pgk-1,2-lacZ* (data not shown).

In situ staining of transformed colonies with X-gal indicated that 10-20% of colonies co-transfected with *Pgk-lacZ* contained blue cells while 60-90% of colonies co-transfected with *Pgk-1,2-lacZ* stained with X-gal. The enhanced expression of β-galactosidase in stably transfected clones containing *Pgk-1,2-lacZ* was similar in co-transfections with three different plasmids carrying the *Pgk-1* promoter driving neo (55), hygromycin (28) or puromycin (58).

Northern blots of RNA isolated from pooled populations of stably transformed cells carrying *Pgk-lacZ* or *Pgk-1,2-lacZ* indicated that the β-galactosidase activities were proportional to the levels of *lacZ* transcript (Figure 2, lanes 1 and 2). This northern blot also indicates that the *Pgk-1,2-lacZ* gene product is spliced to yield an mRNA longer than that from *Pgk-lacZ* due to the presence in the former mRNA of exons I, II, and III derived from the *Pgk-1* gene.

To identify the site(s) in *Pgk-1,2-lacZ* responsible for enhancing its activity in stable transformants we made a number of constructs derived from *Pgk-1,2-lacZ* in which regions of introns 1

and 2 were deleted (constructs *c* to *f*, Figure 1). Some of these constructs deleted exon II. This exon contains 51bp (35) so its removal does not alter the reading frame provided splicing can take place between exons I and III. Transient transfections of the 4 deleted constructs yielded activities between 50% and 125% those of *Pgk-1,2-lacZ* indicating that efficient transcription and splicing did occur in all constructs. In stably transformed cells the β-galactosidase activities from cells carrying constructs *c* to *f* were between 1% and 10% those of *Pgk-1,2-lacZ*. The regions deleted from constructs *d* and e are not overlapping yet both deletions resulted in significant loss of β-galactosidase activity compared to *Pgk-1,2-lacZ* suggesting that at least 2 DNA sequence elements are responsible for the elevated expression of *Pgk-1,2-lacZ* in stable transformants. The β-galactosidase activities in stable clones containing constructs *c* and *f* were lower than those in constructs *d* and *e* consistent with the idea that the β-galactosidase activity in stable transformants might be determined by the overall length of the primary transcript or by the amount of intronic sequence in the construct.

Introns enhance the number of integrated plasmids

The cells in these experiments were exposed to calcium phosphate precipitates containing equal concentrations of DNA and the transient expression data of each was similar, suggesting that each construct was taken up by cells with similar efficiency. However, when slot blots of DNA from pooled populations of stable transformants were probed for *lacZ* sequences we found that the

number of integrated copies of the *lacZ* gene varied greatly in cell populations transfected with different constructs. Figure 3 (lanes 1-6) shows a Southern blot of DNA from pooled populations of cells transformed with constructs a-f shown in Figure 1 and probed for *lacZ* sequences. The intensity of the *lacZ* band from cells transfected with *Pgk-1,2-lacZ* was much higher than for cells transfected with *Pgk-lacZ*.

In 8 of the 15 transfection experiments whose results are reported in Figure 1 the DNA was isolated and analyzed by slot blots probed with *lacZ*. Densitometric traces of these slot blots indicated that in all cases the gene copy numbers assessed by these slot blots were proportional to the β-galactosidase activities measured in the same cell populations. In these 8 experiments, cells transformed with *Pgk-1,2-lacZ* had integrated on average 30 times more *lacZ* sequences than cells tranformed with *Pgk-lacZ*. The number of integrated plasmids per genome varied from experiment-to-experiment. For cells exposed to *Pgk-lacZ* between 4 and 20 *lacZ* genes per genome were present in the pooled population of transformants while cells exposed to *Pgk-1,2-lacZ* integrated between 75 and 600 copies of the *lacZ* plasmid.

Constructs *c*, *d*, *e*, and *f* were integrated at copy numbers lower than *Pgk-1,2-lacZ* (Figure 3, lanes 3-6). Of these 4 deletion constructs, e resulted in the highest expression level in stable transformants (Figure 1) and had the highest average number of integrated copies (Figure 3, lane 5). Construct e also had the smallest internal deletion of these 4 constructs.

Thus, although cells were initially exposed to calcium phosphate precipitates containing equal concentrations of plasmid DNA carrying the *lacZ* gene and although the transient expression studies suggested equal amounts of DNA were taken up by the cells exposed to each construct, the amount of plasmid DNA stably incorporated into high molecular weight genomic DNA varied widely and was proportional to the level of expression of the β-galactosidase encoded by the integrated *lacZ* gene.

Pgk-1 promoter is required for elevated plasmid integration

The elevated copy numbers of integrated *Pgk-1,2-lacZ* plasmid requires intragenic *Pgk-1* sequences. To determine whether the *Pgk-1* promoter was also required for this effect we deleted regions of this promoter from the *Pgk-1,2-lacZ* construct. In construct *h*, Figure 4 the upstream region was reduced from 860bp in *Pgk-1,2-lacZ* to 296bp. The resulting gene had undiminished activity in transient expression assays indicating that the promoter was still active. Construct *h* yielded stable transformants with high β-galactosidase activities (Figure 4) and DNA from these populations of cells contained high copy numbers of the *lacZ* sequence similar to those in cells transfected with *Pgk-1,2-lacZ* (Figure 3, lane 8).

Deletion of the entire promoter to -11bp (construct *i*, Figure 4) yielded a gene with essentially no activity in transient transfection assays. This promoterless construct yielded almost no β-galactosidase activity from stably transformed cells (Figure 4) and the gene copy number was also low, similar to that from *Pgk-lacZ* (data not shown). To determine if another promoter could

substitute for *Pgk-1* we inserted the promoters from Rous sarcoma virus (RSV) (13) and the cytomegalovirus(CMV) (52) into construct *i* upstream of -11bp. The RSV driven construct *j* (Figure 4) was almost as active as *Pgk-1,2-lacZ* in transient assays. Stable transformants with this construct had intermediate β-galactosidase activities and DNA from these stable transformants contain intermediate numbers of the *lacZ* gene (Figure 3, lane 9). The CMV promoter in construct *k* was essentially inactive in P19 cells and stable transformants with this construct contained no β-galactosidase activity and carried few incorporated *lacZ* genes (Figure 3, lane 10).

Thus, it appears that transfected lacZ genes integrate in high copy number only if the construct contains an active promoter, suggesting that transcription through introns 1 and 2 of *Pgk-1* is essential or that elements of the promoter cooperate with intragenic sequences in increasing the gene copy number. To attempt to distinguish between these two possibilities we inserted a DNA fragment into intron 1 which we expected to prematurely truncate transcription driven from the *Pgk-1* promoter. A 460bp PstI-HindIII fragment derived from the last exon of *Pgk-1* contains the site of polyadenylation, the polyA signal and the presumed site of transcription termination (33) and was used to replace a Xho1-HindIII fragment in intron 1 in construct *h*, Figure 1. The β-galactosidase activity in transient assays from this construct were similar to those of *Pgk-1,2-lacZ* indicating that the inserted fragment did not efficiently terminate trascription. Stable

transformants with construct *h* yielded high β-galactosidase activities and these cells contained high numbers of copies of the *lacZ* gene (Figure 3, lane 7).

Our previous work had indicated that the Pst1-HindIII fragment was necessary for efficient expression of *Pgk-1* driven transcripts encoding the selectible neo gene (33); however, transcription was not efficiently terminated by this fragment in intron 1 of construct *h*. Its failure to terminate was probably a result of its insertion into an intron because there is evidence that the 3′ end formation and polyadenylation of transcripts occurs only if the polyadenylation signal is in the 3′ terminal exon (40).

Cell specificity

The transfection experiments described above were done on P19 embryonal carcinoma cells. We have used another line of embryonal carcinoma cells, P10 (32), and noted similar effects to those for P19; that is, the *Pgk-1,2-lacZ* was expressed in stable transformants at 10-30 fold higher levels than was *Pgk-lacZ*. However the behaviour of transfected DNAs was different in cell lines other than embryonal carcinoma. Figure 5 shows results from transfection experiments in which aliquots of the same calcium phosphate DNA precipitate were exposed to P19 and V79 Chinese hamster fibroblasts. The three constructs with active promoters, *Pgk-lacZ*, *Pgk-1,2-lacZ* and construct *c* were all active in V79 cells. Whereas stable transformants of P19 cells contained very different β-galactosidase activities and widely different amounts of integrated plasmid DNA, the stably transformed V79 cells

contained almost equal β-galactosidase activities and similar amounts of the plasmid DNAs (Figure 4B), about 20-30 plasmid copies for genome. Northern blots of RNA from cells stably transfected with *Pgk-lacZ*, and *Pgk-1,2-lacZ* confirmed that the β-galactosidase activities reflect the amount of transcript (Figure 2, lanes 3 and 4). Thus although the number of copies of the integrated plasmid in P19 cells depended on an active promoter and on an intragenic region, these characteristics of the transfected genes did not effect the integrated copy numbers in V79 cells.

Elevated expression of co-transfected genes

In all of the experiments described above the *lacZ* containing constructs were co-transfected along with plasmids encoding selectible genes. In some cases *Pgk-CAT* was also co-transfected as an internal standard. When the DNA prepared from the stably transformed P19 cells was probed for the selected gene or the co-transfected CAT gene we found that these genes were present in numbers proportional to those of *lacZ*. In those transformants with high *lacZ* copy number the copy number of the co-transfected genes were also elevated and expression of these genes was proportionately enhanced.

To investigate the elevated integration of plasmids from co-transfected cells we carried out transfection experiments in which the *Pgk-lacZ* construct was co-precipitated with plasmids carrying the selectible *Pgk-neo* along with another plasmid carrying portions of the *Pgk-1* gene. The first eight exons of the *Pgk-1* are carried on a 17 kbp BamH1 fragment called B17 (1) (Figure 9, construct *2*).

When co-transfected with *Pgk-lacZ* and *Pgk-neo* into P19 cells this B17 fragment enhanced the number of G418 resistant colonies by 2-3 fold and increased the proportion of colonies that contain X-gal staining cells from 20-30% to 60-90% (Figure 6). Co-transfection with B17 had no effect on transient expression from the *Pgk-lacZ* plasmid but resulted in 10-50 fold increases in the β-galactosidase activities from stably transformed cells (Figure 7).

To determine whether enhanced expression of co-transfected plasmids could be achieved with other cloned fragments we co-transfected *Pgk-lacZ* with a 14 kbp EcoR1 fragment containing the human β-actin gene (22) and found that it had essentially no effect on the expression of *Pgk-lacZ* in stable transformants (Figures 6 and 7). Similar results were obtained with cloned fragments containing the mouse *Pgk-2* gene (7), the mouse TIMP gene (12) and a fragment containing the first three exons of the mouse *hprt* gene (48) (data no shown). Co-transfection with total genomic DNA also failed to enhance expression of *Pgk-lacZ*.

Enhanced expression of *Pgk-lacZ* was accompanied by elevated numbers of integrated copies of the transfected plasmid DNA (Figure 8). While co-transfection with the B17 fragment enhanced the numbers of integrated copies, the β-actin gene did not.

To determine whether co-transfection with B17 could enhance the numbers of integrated copies of plasmids carrying genes driven by other promoters, we co-transfected a plasmid carrying the *lacZ* gene driven by the human β-actin promoter (38). This promoter was less than 10% as active as *Pgk-1* but co-transfection with B17

resulted in a 10 fold increase in β-galactosidase expression from the β-actin promoter in stably transformed cells (Figure 7) and was accompanied by a similar increase in the number of copies of the integrated β-actin *lacZ* plasmid (Figure 8). B17 similarly increased expression from reporter genes driven by every other promoter investigated including those of SV40 early gene, the murine sarcoma virus, human retinoblastoma gene and human cardiac actin gene (data not shown).

To investigate the regions of B17 responsible for enhancing the numbers of integrated plasmid copy numbers, we subcloned the 17 kbp fragment into a variety of smaller fragments (Figure 9, constructs *3-7*). None of these significantly enhanced expression of co-transfected *Pgk-lacZ* (Figure 9).

Two internal deletions of B17 were constructed (Figure 9, constructs *8 and 10)*. These two deletions removed non-overlapping regions of B17. Both retained almost as much activity as B17 and enhanced expression of *Pgk-lacZ* by more than 25 fold.

Both regions comprising construct *10* were separately cloned but neither part alone enhanced *Pgk-lacZ* expression (Figure 9, constructs *11* and *12)*. Even when added together, condition 13, these two separately cloned fragments failed to significantly enhance *Pgk-lacZ* expression. Similarly all combinations of constructs *3, 4, 5, and 6* failed to enhance *Pgk-lacZ* expression. Thus it appears that significant enhancement of *Pgk-lacZ* expression in stable transformants occurs only with fragments of B17 that contain the *Pgk-1* promoter along with a region of a significant

length derived from the intragenic portion of *Pgk-1*.

The DNA slot blot shown in Figure 10 indicates that the number of copies of *Pgk-lacZ* are elevated in P19 cells stably co-transfected with those B17 derived constructs able to enhance expression of *Pgk-lacZ*. Both *lacZ* and neo genes were present in elevated copy numbers. B17 had no effect when co-transfected into V79 cells where the copy numbers of integrated *lacZ* and neo genes were uneffected by the co-transfected plasmids.

DISCUSSION

Following calcium phosphate mediated transfection of P19 embryonal carcinoma cells, the number of copies of plasmid DNA integrated into the genomes of recipient cells was 20-40 fold higher when the plasmid contained an active *Pgk-1* promoter driving through an intragenic region derived from the *Pgk-1* gene. Plasmids co-transfected with the *Pgk-1* containing constructs also co-integrated in elevated numbers. The effect of the intragenic region was not to select for those cells containing high copy numbers integrated plasmids because the number of stably transformed cells able to form colonies in selection media was elevated 2-4 fold over the numbers of colonies formed following transfection without the *Pgk-1* derived sequences. The effects of the *Pgk-1* sequences were noted in embryonal carcinoma cell lines but no effect was noted in transfection of fibroblast like cells such as Chinese hamster V79 and NIH3T3 cells.

The magnitude of the increased copy number was roughly

proportional to the length of the intragenic *Pgk-1* region transcribed but did not seem to be dependent on any particular part of the intragenic region. For example, the region between the two Apa1 sites is deleted in construct *c*, Figure 1 and this deletion completely ablates the ability of this gene to be integrated in high copy number. This same Apa1 fragment was deleted from B17 in construct *10*, Figure 9 yet this latter construct remains effective in enhancing the number of integrated copies of the co-transfected *Pgk-1acZ* gene.

The simplest interpretation of our results are that elevated numbers of stably integrated plasmids depends on two factors: (1) the presence of an active promoter which was usually *Pgk-1* in our experiments but could be RSV and (2) the promoter has to drive transcripts that contain at least 6 kbp derived from the *Pgk-1* gene.

Transfected DNA is thought to be taken up by endocytosis (26) but only a small proportion of this DNA enters the nucleus. Plasmid DNA are thought to concatamerize within the transfected cell by homologous and non-homologous recombination (59) followed by a single non-homologous recombination with the host genome to integrate the concatamer. There are at least 3 possible means by which the *Pgk-1* gene might yield elevated gene copy numbers in stably transformed P19 cells. (1) The co-transfected plasmids could be ligated together by homologous recombination more efficiently to yield longer plasmid concatamers prior to integration into the host genome. (2) The *Pgk-1* sequences could

elevate the efficiency with which plasmids undergo non-homologous recombination with the host genome to yield integrated transformants. Genes actively being transcribed are known to be more recombinogenic than inactive genes *in vitro*, in bacteria, in yeast and in mammalian cells (11,18,39,57) consistent with the idea that the *Pgk-1* driven transcript enhances ligation and/or integration. (3) The plasmid containing the *Pgk-1* sequences may carry an origin of DNA replication that amplifies both that plasmid and plasmids with which it concatemerizes. The elevated levels of integrated plasmid DNAs may arise by DNA amplification either before or after integration into the host genome. It has been shown that plasmids bearing mammalian DNA can replicate without integration into the genome (15,19-21,34) and that certain sequences facilitate amplification of DNA sequences following genomic integration (30). The requirement for the *Pgk-1* promoter is consistent with this possibility because transcription factors in yeast (27) and in mammalian viruses (10) appear to play critical roles in the initiation of DNA replication.

The fact that co-transfected plasmids are integrated at elevated numbers in P19 cells strongly suggests that catamerization of co-transfected plasmids does occur; however, it is not clear whether this is the event that is enhanced by the *Pgk-1* sequences. Although the amount of DNA that can be incorporated into a cell varies considerably from one cell line to another (16,36) there has been no previous indication that the amount of DNA integrated into a cell was determined by the nature of the transfected sequences.

A number of genes are associated with DNA sequences that enhance expression of promoters following integration into the genomes of recipient cells and confer levels of expression proportional to gene copy number. These locus control regions, which have been identified flanking the β-globin gene cluster (14), the chick lysozyme gene (56), and the metallothionein gene (44) do not influence the numbers of copies of integrated genes but may contain sites of attachment to the nuclear matrix (56). In pooled populations of P19 cell transfected with plasmids carrying *Pgk-1* sequences, the level of gene expression is proportional to the average number of integrated genes; however, we have not yet investigated individual clones to verify that the expression levels in each clone is copy number dependent.

The effect of *Pgk-1* sequences on plasmid integration was seen in embryonal carcinoma cells and was documented for P19 cells but no effect was evident in V79 fibroblasts. Embryonal carcinoma cells are derived from and resemble pluripotential embryonic cells and/or primordial germ cells. Embryonal carcinoma cells differ from fibroblasts in a number of ways. For example, the transcription factors present are distinct with embryonal carcinoma cells expressing certain transcription factors such as Oct3/4 (42,51) and PEA3 (60) not normally expressed in somatic cells. A number of anti-oncogenes are also present at unusual levels in EC cells. For example, the retinoblastoma susceptibility gene is expressed at very low levels in EC cells (53) while the p53 gene is expressed at very high levels in this cell type (24). This latter

gene may be pertinent because its loss from tumors is associated with acquired chromosome instability and gene amplification (25,61). Perhaps the germ cell nature of EC cells confers on them unusual capacities for DNA repair and mechanisms for dealing with naked transfected DNA.

Although we have investigated a number of other genes, only the *Pgk-1* gene appears to enhance the numbers of stably integrated plasmids in P19 cells. This peculiar property of *Pgk-1* may be related to the fact that this gene is normally X-linked and subject to X-chromosome inactivation. Alternatively, the expression of the *Pgk-1* promoter in embryonic cells is particularly robust (33) and a high density of transcription complexes on transfected genes in P19 cells may make them very recombinogenic.

Regardless of the mechanism by which *Pgk-1* sequences confer elevated numbers of plasmids in stable transformants it is useful to be able to achieve a high level of co-transfection in P19 cultures. Because co-transfected plasmids are also integrated at elevated copy numbers regardless of the sequences cloned into those plasmids, we can use B17 to increase the output from these co-transfected genes. In transfection experiments utilizing regulated promoters such as those of the human cardiac actin gene, the mouse Hox1.6 and RARβ genes and the mouse retinoblastoma susceptibility gene, the promoters remain regulated even when integrated at a relatively high copy number (45,47,53). Thus by co-transfecting with B17 we can investigate even relatively weak promoters by enhancing the overall signal without compromising its regulation.

ACKNOWLEDGEMENTS

The authors thank Dr. John Bell for discussions. This work was supported by the Medical Research Council of Canada. MWM is a Terry Fox Scientist of the Canadian NCI and SF was supported by a post-doctoral fellowship from the MRCC.

FIGURE LEGENDS

Figure 1. Intragenic regions of *Pgk-1* enhance expression from stably transformed P19 cells. The left panel shows schematic representations of the expression constructs used for transfections. The *E. coli lacZ* coding region is shown as the striped bar, the *Pgk-1* coding regions are indicated as black bars, and *Pgk-1* non-coding regions are shown as open bars. The restriction map below shows exons I, II, III and XI derived from the *Pgk-1* gene and the restriction sites used for the constructs (B, BamH1; E, EcoR1; X, Xho1; H, HindIII; A, Apa1; P, Pst1; N, Nae1). In construct *g* the region in intron 1 between the Xho1 and HindIII sites was replaced with the Pst1-HindIII fragment derived from exon XI. The right panel shows the β-galactosidase activities normalized to that obtained from *Pgk-1,2-lacZ,* construct *b*. The open bars are from transient expression studies in which cells were harvested 48hrs following transfection. The black bars are from cells selected for 8 days in G418, hygromycin, or puromycin. Error bars represent standard errors. The results are derived from 15 transfection experiments in which there are 6-21 independent measures of the activity from each of the constructs. At least 2 independent DNA preparations from each construct are represented.

Figure 2. *LacZ*-containing transcripts in P19 and V79 cells transformed with *Pgk-lacZ* (construct *a*) and *Pgk-1,2-lacZ* (construct *b*). RNA was isolated from pools of cells stably transformed with

Pgk-neo and constructs *a* and *b* (Figure 1) and northern blots of this RNA were probed for *lacZ* sequences. The arrow indicates the *lacZ* transcript. Those transcripts derived from *Pgk-1,2-lacZ* migrate slightly more slowly because they contain exons I, II, and III from *Pgk-1*. The absence of slowly migrating bands from cells carrying *Pgk-1,2-lacZ* indicates efficient splicing of the primary transcript.

Figure 3. The *lacZ* sequences are incorporated into stably transformed cells with variable copy number. Cells were co-transfected with *Pgk-puro* along with one of the constructs shown in Figures 1 or 4. Stable transformants were selected in puromycin and DNA extracted from the pooled population of colonies. This DNA was digested with EcoR1, electrophoresed through agarose, blotted and probed for lacZ sequences. EcoR1 cuts twice in all constructs releasing *lacZ*-containing fragments which can vary in size with the construct. Lanes 1 and 2 were transfected with *Pgk-lacZ* and *Pgk-1,2-lacZ* respectively. Lanes 1-7 were from cells transfected with constructs *a-g* shown in Figure 1. Lanes 8-10 carry constructs *h-k* shown in Figure 4.

Figure 4. Modifications to the promoter of *Pgk-1,2-lacZ* effect plasmid copy numbers in stable transformants. Constructs *a* and *b* are *Pgk-lacZ* and *Pgk-1,2-lacZ* as shown in Figure 1. The arrow indicates the transcription start site. Constructs *h, i, j* and *k* were derived from *Pgk-1,2-lacZ* by modification of the region

upstream of the transcription start site. Construct *h* contains an upstream region reduced to 296bp. Construct *i* carries only 11bp upstream of the transcription start site and has no promoter activity. Construct *j* is derived from *i* by the insertion of a 520bp fragment containing the RSV promoter (13). Construct *k* is derived from *i* and carries a 700bp fragment containing the CMV promoter (52). Expression of β-galactosidase in transient (open bars) and in stable transformants (black bars) is indicated on the right following normalization to the expression from *Pgk-1,2-lacZ*. The results shown are the averages of between 3 and 8 independent experiments for each construct.

Figure 5. Transformation of fibroblasts is not enhanced by *Pgk-1* intragenic regions. DNA-calcium phosphate co-precipitates containing constructs *a*, *b* and *c* (Figure 1) and *i* (Figure 4) were prepared and aliquots of the same preparation transfected into P19 embryonal carcinoma and V79 fibroblast cells. The activities in transient (open bars) and stable transformants (black bars) were measured and normalized to those of *Pgk-1,2-lacZ*. The levels of expression in the 2 cell lines were similar. The results shown are the average from 3 experiments. Panel B. DNA from stable transformants were isolated and aliquots of 1μg were immobilized on membranes using a slot blot apparatus. These membranes were probed for *lacZ* sequences. Signals of roughly equal intensity were present in V79 cells transformed with all 4 constructs while the signals from P19 populations were much higher from cells carrying

the *Pgk-1,2-lacZ* plasmid.

Figure 6. B17 enhances expression from a co-transfected plasmid carrying *Pgk-lacZ*. P19 cells were transfected with a mixture of 3 plasmid DNAs. Each mixture contained *Pgk-lacZ* and *Pgk-neo* along with either pGEM4, B17, or *β-actin*. 10^5 cells were plated in medium containing G418 and dishes harvested 8 days later. Colonies were fixed and stained with X-gal. More than a thousand colonies formed on each dish but only those that are X-gal stained are apparent in the photographs.

Figure 7. B17 enhances expression from promoters in co-transfected plasmids. P19 cells were co-transfected with 4μg of each of 3 plasmids: *Pgk-neo*, either *Pgk-lacZ* or *β-actin-lacZ*, and either pGEM4, B17, or *β-actin*. The β-galactosidase activities were normalized to those of *Pgk-lacZ* (upper group of 3) or *β-actin-lacZ* (lower group of 3) co-transfected with B17 and the results recorded from both transient and stable transfection experiments. Results are averaged from 4 experiments. The absolute activity from *β-actin-lacZ* was less than 10% that of *Pgk-lacZ* in these experiments.

Figure 8. Southern blot analysis of DNA from stable transformants. DNA was isolated from cells selected in G418 following transfection with *Pgk-neo*, *Pgk-lacZ* or *β-actin-lacZ*, and pGEM4, B17, or *β-actin*. Genomic DNA was digested with EcoR1, electrophoresed, blotted, and probed for *lacZ* sequences. 10 μg of genomic DNA was loaded onto

each lane.

Figure 9. Enhanced expression of co-transfected plasmids requires select regions of B17. The diagrams on the left represent those regions of B17 used in co-transfection experiments. The restriction map shows the first 8 exons of *Pgk-1* {786} along with the sites used for the constructs below. Construct *2* is B17, constructs *8* and *10* carry internal deletions. Number *13* is a condition in which two separately cloned fragments were co-transfected together. The β-galactosidase activities reported on the right are from stably transformed colonies formed in G418 following co-transfection of cells with DNA mixtures of 4μg each of *Pgk-lacZ*, *Pgk-neo* and the separately cloned fragments indicated on the left. These activities are normalized to that obtained with vector sequences alone, condition *1*. These results represent the averages from between 3 and 14 experiments for each construct. The error bars on lanes *2, 8 and 10* represent the standard errors. Error bars are not placed on the other conditions because the bars are too small to be seen.

Figure 10. Slot blot analysis of DNA from stable transformants of P19 and V79 cells co-transfected with *Pgk-lacZ* and B17. DNA was isolated from stable transformants transfected with *Pgk-lacZ* along with plasmid vector, B17, construct *8*, or construct *9*. 10 and 5μg of DNA were loaded onto adjacent pairs of slots and probed for *lacZ* and *neo* sequences as indicated.

REFERENCES

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

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

3. **Aronow, B., D. Lattier, R. Silbiger, M. Dusing, J. Hutton, G. Jones, J. Stock, J. McNeish, S. Potter, D. Witte, and D. Wiginton.** 1989. Evidence for a complex regulatory array in the first intron of the human adenosine deaminase gene. Genes Dev. 3:1384-1400.

4. **Aronow, B.J., R. N. Silbiger, M. R. Dusing, J. L. Stock, K. L. Yager, S. S. Potter, J. J. Hutton, and D. A. Wiginton.** 1992. Functional analysis of the human adenosine deaminase gene thymic regulatory region and its ability to generate position- independent transgene expression. Mol. Cell. Biol. **12**:4170-4185.

5. **Ash, J., Y. Ke, M. Korb, and L. F. Johnson.** 1993. Introns are essential for growth-regulated expression of the mouse thymidylate synthase gene. Mol. Cell. Biol. **13**:1565-1571.

6. **Auffray, C. and F. Rougeon.** 1980. Purification of mouse

immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. 107:303-314.

7. Boer, P.H., C. N. Adra, Y. F. Lau, and M. W. McBurney. 1988. The testis-specific phosphoglycerate kinase gene, pgk-2, is a recruited retroposon. Mol. Cell. Biol. 7:3107-3112.

8. Boer, P.H., H. Potten, C. N. Adra, K. Jardine, G. Mullhofer, and M. W. McBurney. 1990. Polymorphisms in the coding and non-coding regions of murine pgk-1 alleles. Biochem. Genet. 28:299-308.

9. Chen, C. and H. Okayama. 1987. High efficiency transformation of mammalian cells by plasmid DNA. Mol. Cell. Biol. 7:2745-2752.

10. DePamphilis, M.L. 1993. Eukaryotic DNA replication: Anatomy of an origin. Annu. Rev. Biochem. 62:29-63.

11. Dröge, P. 1993. Transcription-driven site-specific DNA recombination *in vitro*. Proc. Natl. Acad. Sci. USA 90:2759-2763.

12. Gewert, D.R., B. Coulombe, M. Castelino, D. Skup, and B. R. G. Williams. 1987. Characterization and expression of a murine gene homologous to human EPA/TIMP: a virus-induced gene in the mouse. EMBO J. 6:651-657.

13. **Gorman, C.M., G. T. Merlino, M. C. Willingham, I. Pastan, and B. H. Howard.** 1982. The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. Proc. Natl. Acad. Sci. USA 79:6777-6781.

14. **Grosveld, F., G. Blom van Assendelft, D. R. Greaves, and G. Kollias.** 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell 51:975-985.

15. **Heinzel, S.S., P. J. Krysan, C. T. Tran, and M. P. Calos.** 1991. Autonomous DNA replication in human cells is affected by the size and the source of the DNA. Mol. Cell. Biol. 11:2263-2272.

16. **Hoeijmakers, J.H.J, H. Odijk, and A. Westerveld.** 1987. Differences between rodent and human cell lines in the amount of integrated DNA after transfection. Exp. Cell Res. 169:111-119.

17. **Hurt, M.M., T. L. Bowman, and W. F. Marzluff.** 1991. A common transcriptional activator is located in the coding region of two replication-dependent mouse histone genes. Mol. Cell. Biol. 11:2929-2936.

18. **Kohwi, Y. and Y. Panchenko.** 1993. Transcription-dependent recombination induced by triple- helix formation. Genes Dev. 7:1766-1778.

19. **Krysan, P.J. and M. P. Calos.** 1991. Replication initiates at multiple locations on an autonomously replicating plasmid in human cells. Mol. Cell. Biol. 11:1464-1472.

20. **Krysan, P.J., J. G. Smith, and M. P. Calos.** 1993. Autonomous replication in human cells of multimers of specific human and bacterial DNA sequences. Mol. Cell. Biol. 13:2688-2696.

21. **Landry, S. and M. Zannis-Hadjopoulos.** 1991. Classes of autonomously replicating sequences are found among early-replicating monkey DNA. Biochem. Biophys. Acta 1088:234-244.

22. **Leavitt, J., P. Gunning, P. Porreca, S-Y. Ng, C-S. Lin, and L. Kedes.** 1984. Molecular cloning and characterization of mutant and wilt-type human beta-actin genes. Mol. Cell. Biol. 4:1961-1969.

23. **Lin, H., K. E. Yutzey, and S. F. Konieczny.** 1991. Muscle-specific expression of the troponin I gene requires interactions between helix-loop-helix muscle regulatory factors and ubiquitous transcription factors. Mol. Cell. Biol. 11:267-280.

24. **Linzer, D.I.H. and A. J. Levine.** 1979. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. Cell 17:43-52.

25. **Livingstone, L.R., A. White, J. Sprouse, E. Livanos, T. Jacks, and T. D. Tlsty.** 1992. Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53. Cell 70:923-935.

26. **Loyter, A., G. Scangos, and F. H. Ruddle.** 1982. Mechanisms of DNA uptake by mammalian cells: fate of exogenously added DNA monitored by the use of fluorescent dyes. Proc. Natl. Acad. Sci. USA 79:422-426.

27. **Lue, N.F. and R. D. Kornberg.** 1993. A possible role for the yeast TATA-element-binding protein in DNA replication. Proc. Natl. Acad. Sci. USA 90:8018-8022.

28. **Lupton, S.D., L. L. Brunton, V. A. Kalberg, and R. W. Overell.** 1991. Dominant positive and negative selection using a hygromycin phosphotransferase-thymidine kinase fusion gene. Mol. Cell. Biol. 11:3374-3378.

29. **Maniatis, T., T. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

30. **McArthur, J.G. and C. P. Stanners.** 1991. A genetic element which increases the frequency of gene amplification. J. Biol. Chem. 266:6000-6005.

31. McBurney, M.W. and B. J. Rogers. 1982. Isolation of male murine embryonal carcinoma cells and their chromosome replication patterns. Dev. Biol. 89:503-508.

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

33. McBurney, M.W., L. C. Sutherland, C. N. Adra, B. Leclair, M. A. Rudnicki, and K. Jardine. 1991. The mouse Pgk-1 gene promoter contains an upstream activator sequence. Nucleic Acids Res. 19:5755-5761.

34. McWhinney, C. and M. Leffak. 1990. Autonomous replication of a DNA fragment containing the chromosomal replication origin of the human c-myc gene. Nucleic Acids Res. 18:1233-1242.

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

36. Monroe, T.J., M. C. Muhlmann-Diaz, M. J. Kovach, J. O. Carlson, J. S. Bedford, and B. J. Beaty. 1992. Stable transformation of a mosquito cell line results in extraordinarily high copy numbers of

the plasmid. Proc. Natl. Acad. Sci. USA 89:5725-5729.

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

38. **Ng, S-Y., P. Gunning, R. Eddy, P. Ponte, J. Leavitt, T. Shows, and L. Kedes.** 1985. Evolution of the functional human beta-actin gene and its multi-pseudogene family: conservation of noncoding regions and chromosomal dispersion of pseudogenes. Mol. Cell. Biol. 5:2720-2732.

39. **Nickoloff, J.A.** 1992. Transcription enhances intrachromosomal homologous recombination in mammalian cells. Mol. Cell. Biol. 12:5311-5318.

40. **Niwa, M., C. C. MacDonald, and S. M. Berget.** 1992. Are vertebrate exons scanned during splice-site selection. Nature 360:277-280.

41. **Norton, P.A. and J. M. Coffin.** 1985. Bacterial beta-gal as a marker of Rous Sarcoma Virus gene expression and replication. Mol. Cell. Biol. 5:281-290.

42. **Okamoto, K., H. Okazawa, A. Okuda, M. Sakai, M. Muramatsu, and**

H. Hamada. 1990. A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell 60:461-472.

43. **Ottavio, L., C. -D. Chang, M. -G. Rizzo, S. Travali, C. Casadevall, and R. Baserga.** 1990. Importance of introns in the growth regulation of mRNA levels of the proliferating cell nuclear antigen gene. Mol. Cell. Biol. 10:303-309.

44. **Palmiter, R.D., E. P. Sandgren, D. M. Koeller, and R. L. Brinster.** 1993. Distal regulatory elements from the mouse metallothionein locus stimulate gene expression in trransgenic mice. Mol. Cell. Biol. 13:5266-5275.

45. **Pari, G., K. Jardine, and M. W. McBurney.** 1991. Multiple CArG boxes in the human cardiac actin gene promoter required for expression in embryonic cardiac muscle cells developing in vitro from embryonal carcinoma cells. Mol. Cell. Biol. 11:4796-4803.

46. **Pearson, B., P. L. Wolf, and J. A. Vazquez.** 1963. A comparative study of new indolyl compounds to localize beta-gal in tissues. Lab. Invest. 12:1249-1259.

47. **Pratt, M.A.C., A. W. Langston, L. J. Gudas, and M. W. McBurney.** 1993. Retinoic acid fails to induce expression of Hox genes in differentiation-defective murine embryonal carcinoma cells carrying

a mutant gene for alpha retinoic acid receptor. Differentiation 53:105-113.

48. **Reid, L.H., R. G. Gregg, O. Smithies, and B. H. Koller.** 1990. Regulatory elements in the introns of the human *HPRT* gene are necessary for its expression in embryonic stem cells. Proc. Natl. Acad. Sci. USA 87:4299-4303.

49. **Rudnicki, M.A. and M. W. McBurney.** 1987. Cell culture methods and induction of differentiation of embryonal carcinoma cell lines. In Teratocarcinomas and embryonic stem cells, a practical approach. E. J. Robertson, editor. IRL Press, Oxford. 19-49.

50. **Sap, J., L. de Magistris, H. Stunnenberg, and B. Vennstrom.** 1990. A major thyroid hormone response element in the third intron of the rat growth hormone gene. EMBO J. 9:887-896.

51. **Schöler, H.R., S. Ruppert, N. Suzuki, K. Chowdhury, and P. Gruss.** 1990. New type of POU domain in germ line-specific protein Oct- 4. Nature 344:435-439.

52. **Seed, B.** 1987. An LFA-3 cDNA encodes a phospholipid-linked membrane protein homologous to its receptor CD2. Nature 329:840-842.

53. **Slack, R.S., P. A. Hamel, T. S. Bladon, R. M. Gill, and M. W.**

McBurney. 1993. Regulated expression of the retinoblastoma gene in differentiating embyronal carcinoma cells. Oncogene 8:1585-1591.

54. Soriano, P., G. Friedrich, and P. Lawinger. 1991. Promoter interactions in retrovirus vectors introduced into fibroblasts and embrycnic stem cells. J. Virol. 65:2314-2319.

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

56. Stief, A., D. M. Winter, W. H. Strätling, and A. E. Sippel. 1989. A nuclear DNA attachment element mediates elevated and position- independent gene activity. Nature 341:343-345.

57. Thomas, B.J. and R. Rothstein. 1989. Elevated recombination rates in transcriptionally active DNA. Cell 56:619-630.

58. Vara, J.A., A. Portela, J. Ortin, and A. Jimenez. 1986. Expression in mammalian cells of a gene from Streptomyces alboniger conferring puromycin resistance. Nucleic Acids Res. 14:4617-4624.

59. Wigler, M., R. Sweet, G. K. Sim, B. Wold, A. Pellicer, E. Lacy, T. Maniatis, S. Silverstein, and R. Axel. 1977. Transformation of mammalian cells with genes from procaryotes and eucaryotes. Cell 16:777-785.

60. **Xin, J.-H., A. Cowie, P. Lachance, and J. A. Hassell.** 1992. Molecular cloning and characterization of PEA3, a new member of the *ets* oncogene family that is differentially expressed in mouse embryonic cells. Genes Dev. 6:481-496.

61. **Yin, Y., M. A. Tainsky, F. Z. Bischoff, L. C. Strong, and G. M. Wahl.** 1992. Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles. Cell 70:937-948.

FIG1

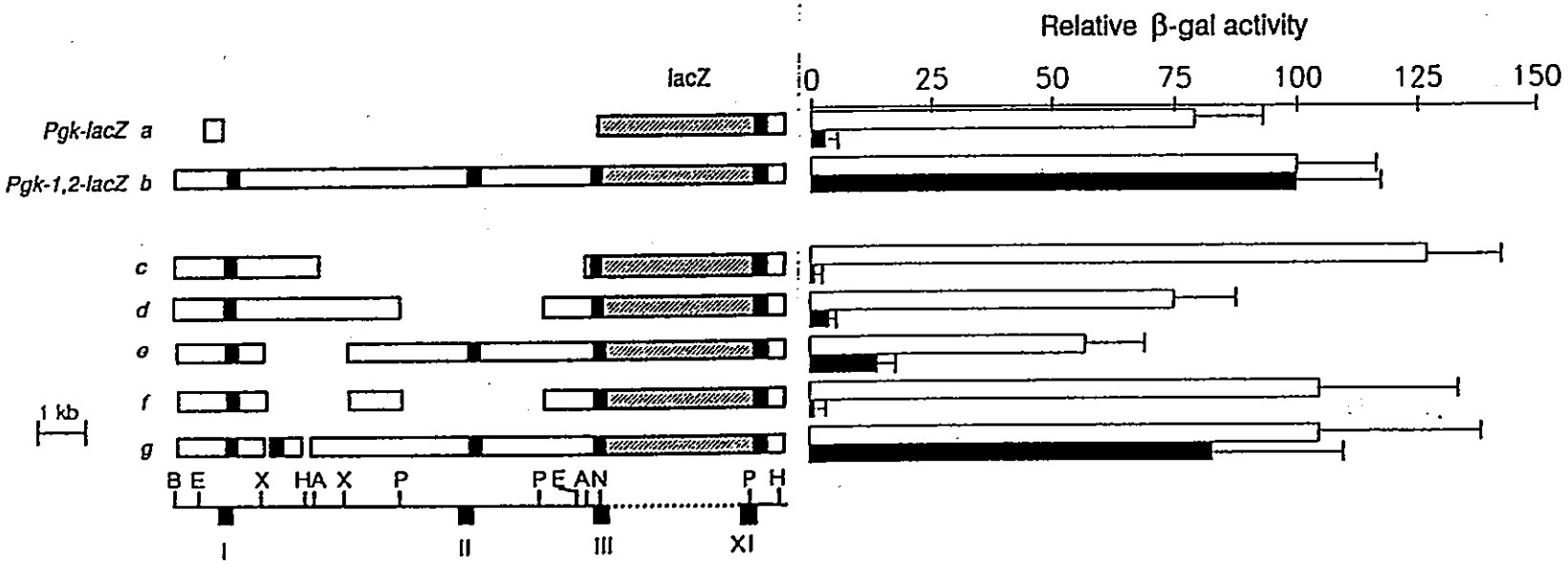

Relative β-gal activity
lacZ
0
25
50
75
100
125
150
Pgk-lacZ a
Pgk-1,2-lacZ b
c
d
e
f
g
1 kb
B E X HA X P P E AN P H
I
II
III
XI

Fig 2

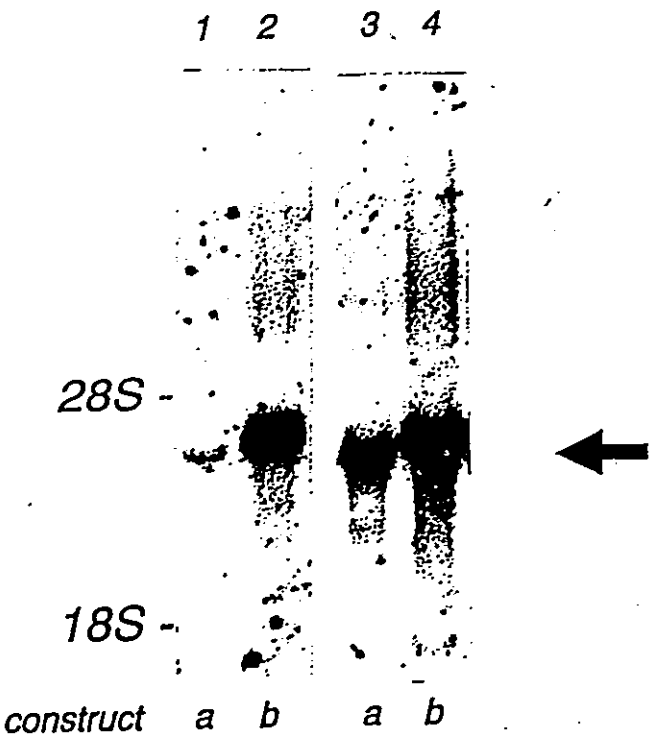


FIG 3

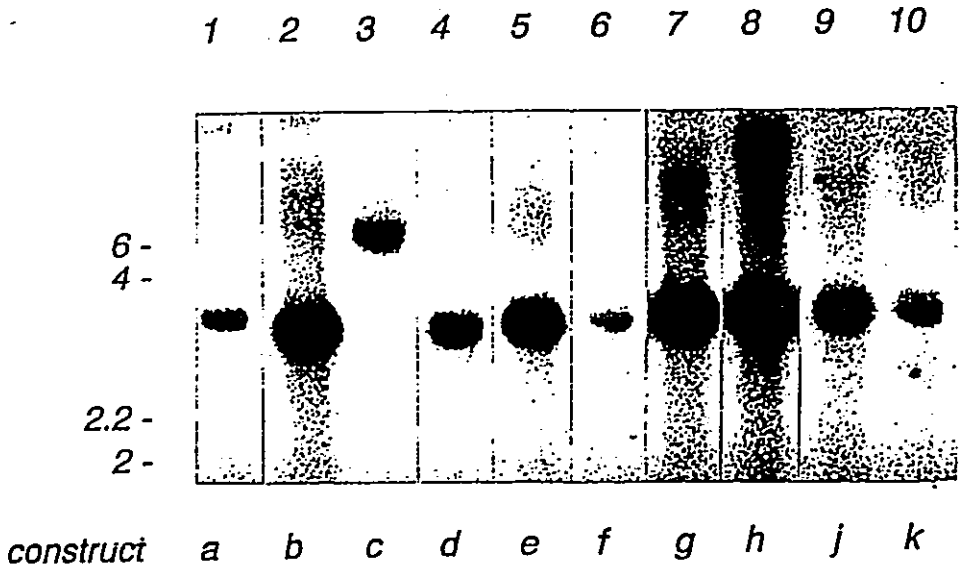
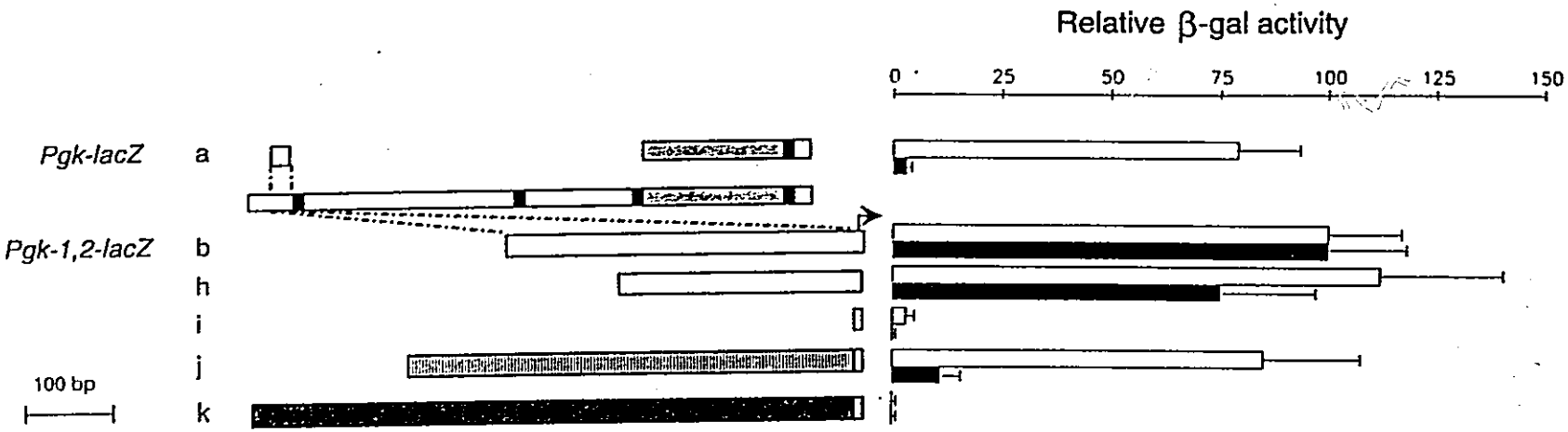


FIG 4


Relative β-gal activity
0
25
50
75
100
125
150
Pgk-lacZ
a
Pgk-1,2-lacZ
b
h
i
j
k
100 bp

F.S 5A

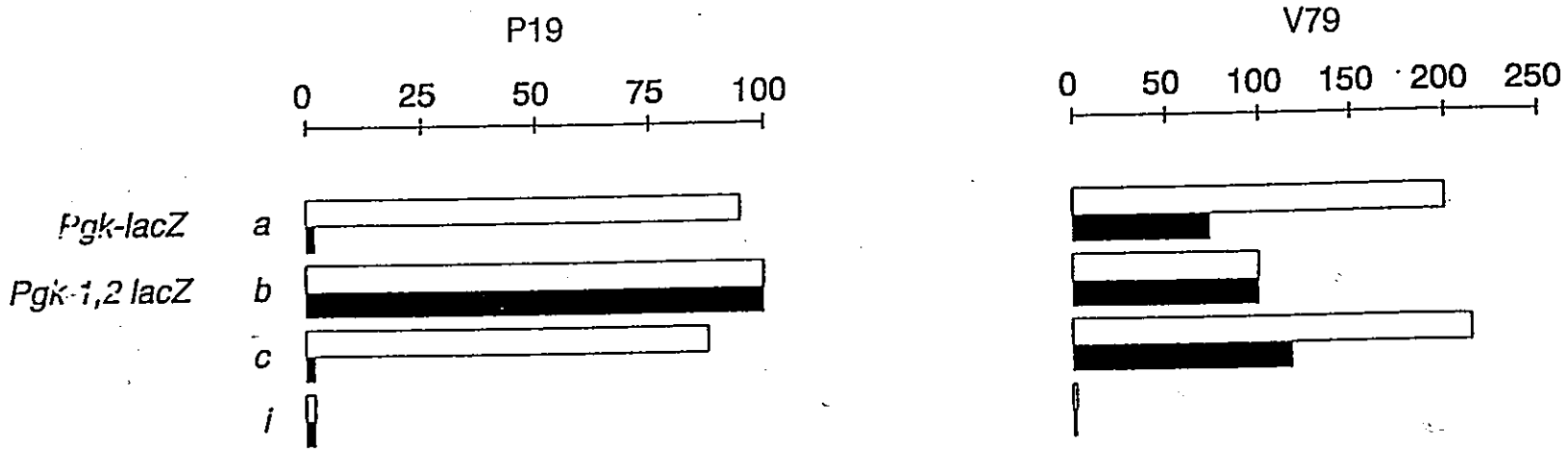

P19
0
25
50
75
100
Pgk-lacZ
a
Pgk-1,2 lacZ
b
c
i
V79
0
50
100
150
200
250

FIG 5B

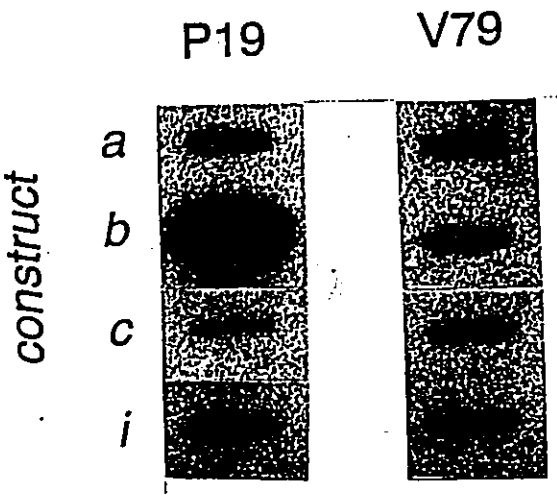
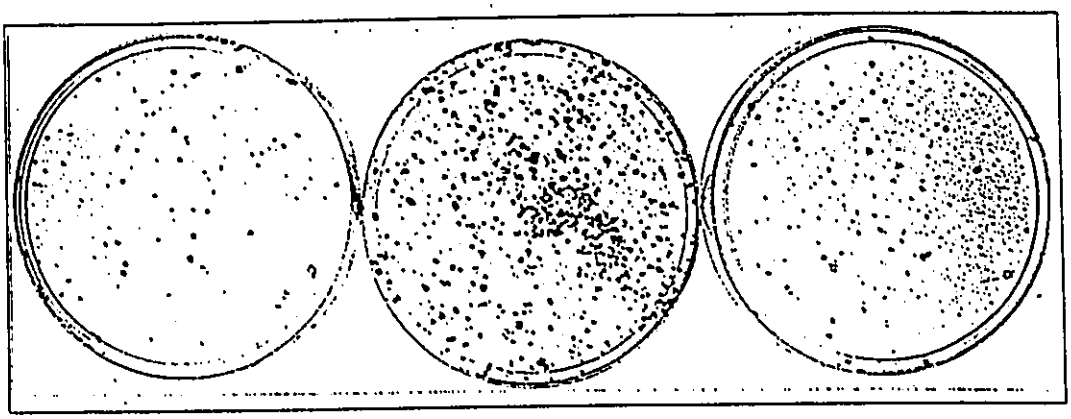


FIG 6



pGEM4 B17 β-*actin*

FIG 1

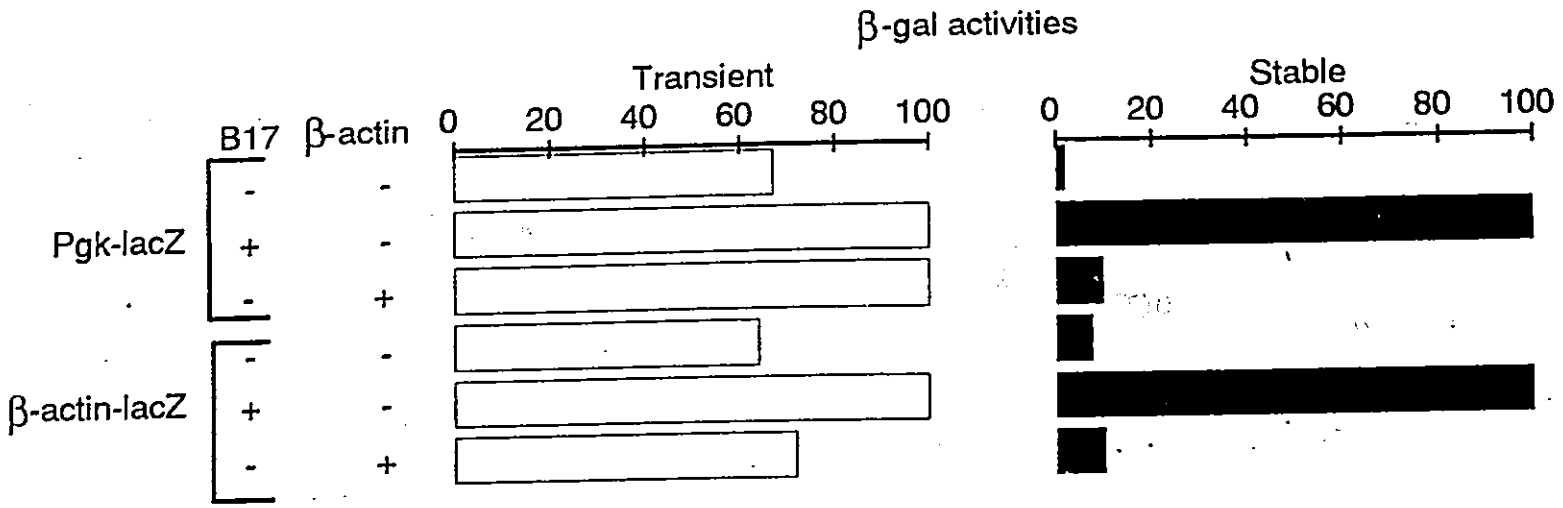
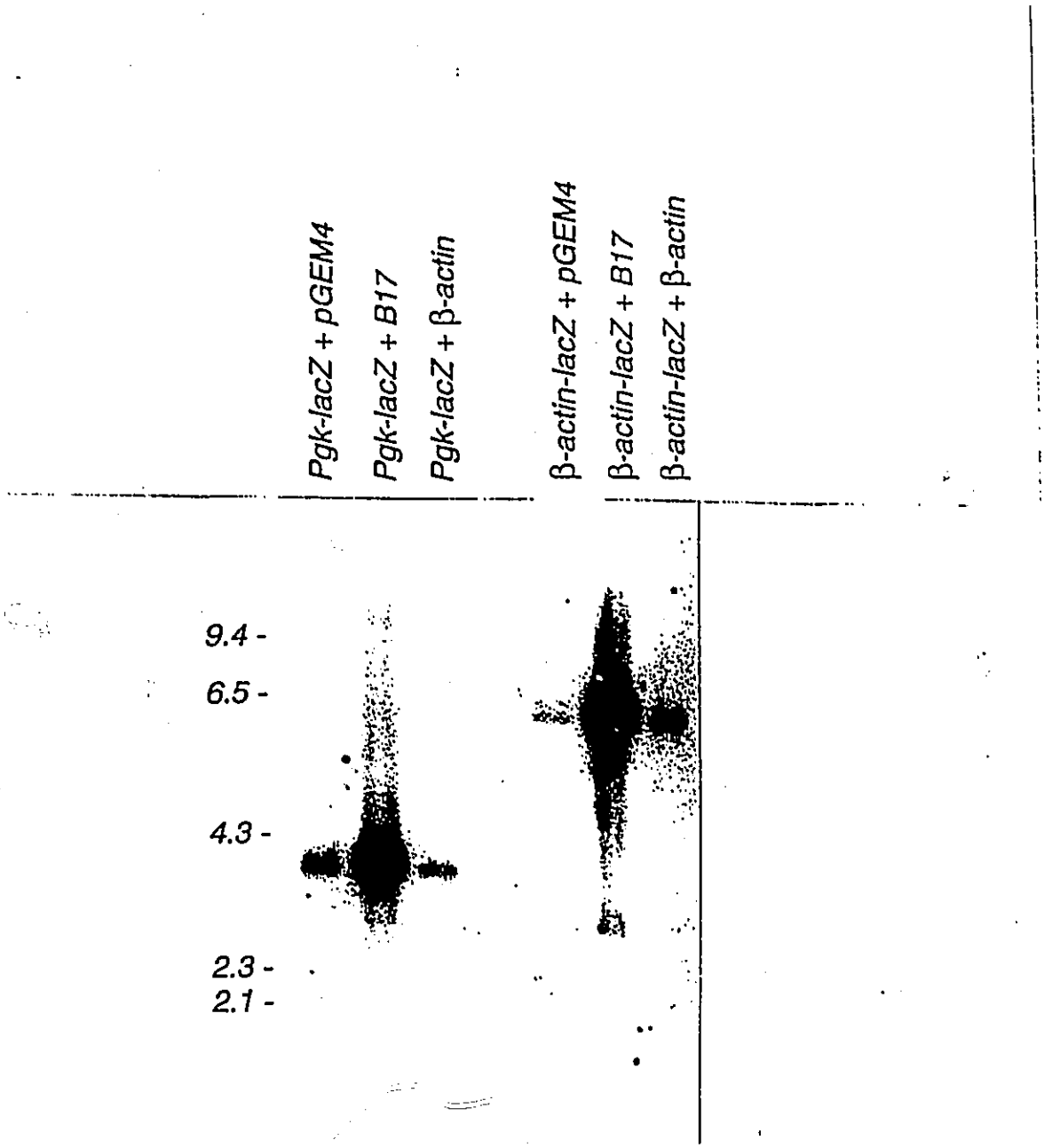

β-gal activities
Transient
Stable
0
20
40
60
80
100
B17
β-actin
Pgk-lacZ
β-actin-lacZ
-
+

FIG 8


Pgk-lacZ + pGEM4
Pgk-lacZ + B17
Pgk-lacZ + β-actin
β-actin-lacZ + pGEM4
β-actin-lacZ + B17
β-actin-lacZ + β-actin
9.4 -
6.5 -
4.3 -
2.3 -
2.1 -

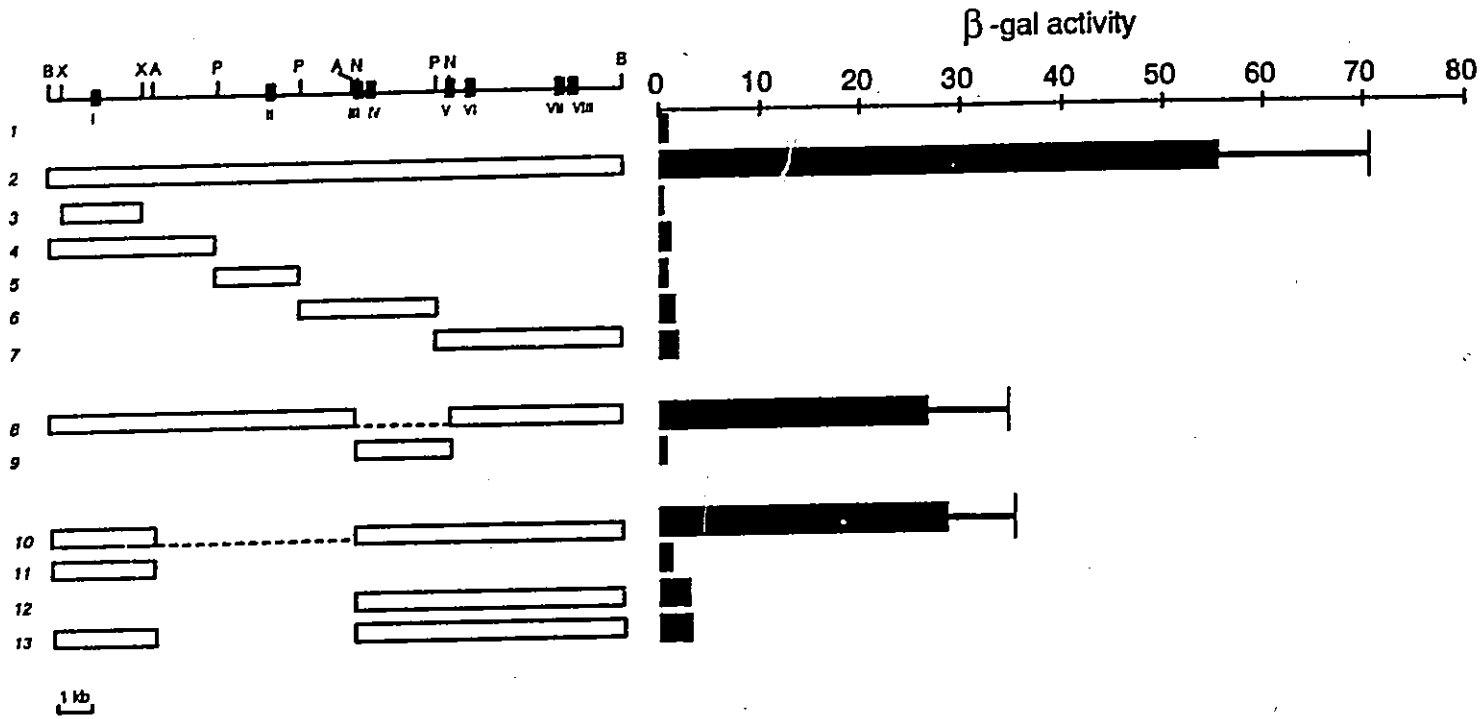
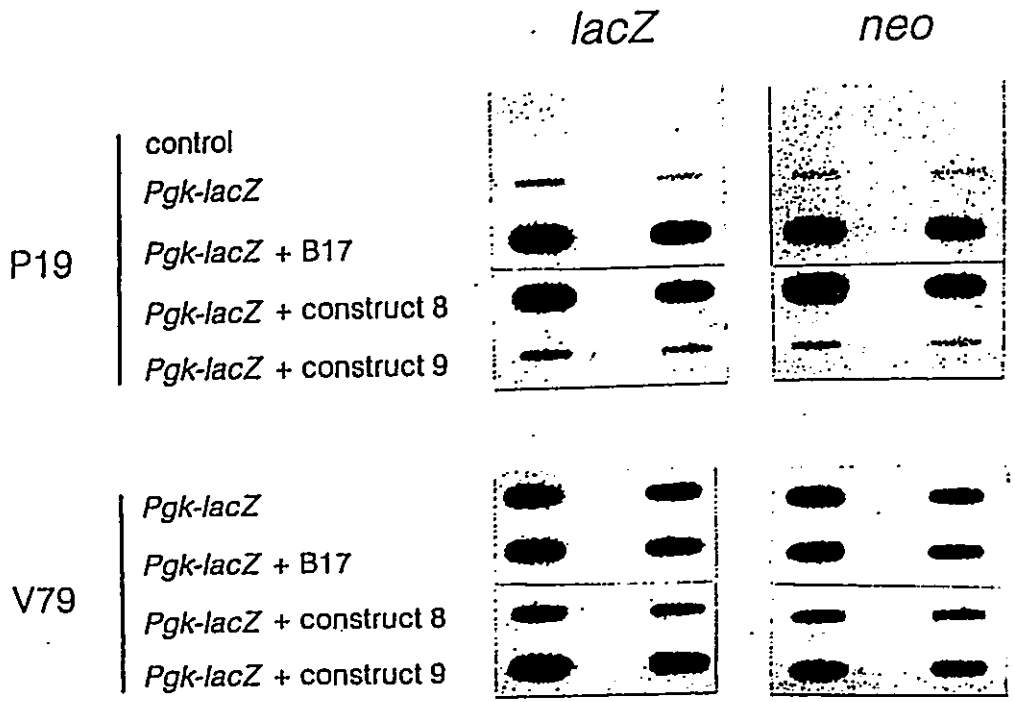

β -gal activity
BX
XA
P
P
A N
P N
B
I
II
III IV
V VI
VII VIII
0
10
20
30
40
50
60
70
80
1
2
3
4
5
6
7
8
9
10
11
12
13
1 kb

FIG 10


lacZ
neo
P19
control
Pgk-lacZ
Pgk-lacZ + B17
Pgk-lacZ + construct 8
Pgk-lacZ + construct 9
V79
Pgk-lacZ
Pgk-lacZ + B17
Pgk-lacZ + construct 8
Pgk-lacZ + construct 9

REFERENCES

1. Abel, T. and Maniatis, T. (1989) Gene regulation:action of leucine zippers. nature 341:24-25.

2. Adachi, Y., Käs, E. and Laemmli, U.K. (1989) Preferential, cooperative binding of DNA topoisomerase II to scaffold-associated regions. EMBO J. 8:3997-4006.

3. Adams, C.C. and Workman, J.L. (1993) Nucleosome displacement in transcription. Cell 72:305-308.

4. Adra, C.N., Ellis, N.E. and McBurney, M.W. (1988) Family of mouse phosphoglycerate kinase genes and pseudogenes. Som.Cell and Mol.Gen. 14:69-81.

5. Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshit, S., Nishio, Y., Nakajima, T., Hirano, T. and Kishimot, T. (1990) A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. EMBO J. 9:1897-1906.

6. Archer, T.K., Lefebvre, P., Wolford, R.G. and Hager, G.L. (1992) Transcription factor loading on the MMTV promoter:a bimodal mechanism for promoter activation. Science 255:1573-1576.

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

8. Auwerx, J. and Sassone-Corsi, P. (1991) IP-1:a dominant inhibitor of Fos/Jun whose activity is modulated by phosphorylation. Cell 64:983-993.

9. Ballabio, A. and Willard, H.F. (1992) Mammalian X-chromosome inactivation and the XIST gene. Curr.Op.Genet.Devel. 2:439-447.

10. Baniahmad, A., Ha, I., Reinberg, D., Tsai, S., Tsai, M.-J. and O'Malley, B. (1993) Interaction of human thyroid hormone receptor β with transcription factor TFIIB may mediate target gene derepression and activation by thyroid hormone. Proc.Natl.Acad.Sci.USA 90:8832-8836.

11. Barinaga, M. (1991) Dimers direct development. Science 251:1176-1177.

12. Beckmann, H. and Kadesch, T. (1991) The leucine zipper of TFE3 dictates helix-loop-helix dimerization specificity. Genes Dev. 5:1057-1066.

13. Beckmann, H., Su, L.K. and Kadesch, T. (1990) TFE3:a helix-loop-helix zipper protein that activates transcription through the immunoglobulin enhancer mu-E3 motif. Genes Dev. 4:167-179.

14. Ben-David, Y., Giddens, E.B., Letwin, K. and Bernstein, A. (1991) Erythroleukemia induction by Friend murine leukemia virus:insertional activation of a new member of an ets gene family, Fli-1, closely linked to c-ets-1. Genes Dev. 5:908-918.

15. Benezra, R., Davis, R.L., Lassar, A., Tapscott, S., Thayer, M., Lockshon, D. and Weintraub, H. (1990) Id:a negative regulator of helix-loop-helix DNA-binding proteins. Ann.NY.Acad. 599:1-11.

16. Bengal, E., Ransone, L., Scharfmann, R., Dwarki, V.J., Tapscott, S.J., Weintraub, H. and Verma, I.M. (1992) Functional antagonism between c-Jun and MyoD proteins:a direct physical association. Cell 68:507-519.

17. Berrios, M., Osheroff, N. and Fisher, P.A. (1985) In situ localization of DNA topoisomerase II, a major polypeptide component of the Drosophila nuclear matrix fraction. Proc.Natl.Acad.Sci.USA 82:4142-4146.

18. Bhat, N.K., Fisher, R.J., Fujiwara, S., Ascione, R. and Papas, T.S. (1987) Temporal and tissue-specific expression of mouse ets genes. Proc.Natl.Acad.Sci.USA 84:3161-3165.

19. Bhat, N.K., Thompson, C.B., Lindsten, T., June, C.H., Fujiwara, S., Koizumi, S., Fisher, R.J. and Papas, T.S. (1990) Reciprocal expression of human ETS1 and ETS2 genes during T-cell activation:regulatory role for the protooncogene ETS1. Proc.Natl.Acad.Sci.USA 87:3723-3727.

20. Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl.AcidsRes. 7:1513-1520.

21. Blackwell, T.K. and Weintraub, H. (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science 250:1104-1110.

22. Blackwood, E.M. and Eisenman, R.N. (1991) Max:a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with myc. Science 251:1211-1217.

23. Blasquez, V.C., Xu, M., Moses, S.C. and Garrard, W.T. (1989) Immunoglobulin kappa gene expression after stable integration.I. Role of the intronic MAR and enhancer in plasmacytoma cells. J.Biol.Chem. 264:21183-21189.

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

25. Boer, P.H., Potten, H., Adra, C.N., Jardine, K., Mullhofer, G. and McBurney, M.W. (1990) Polymorphisms in the coding and noncoding regions of murine Pgk-1 alleles. Biochem.Genet. 28:299-308.

26. Boulikas, T. (1993) Nature of DNA sequences at the attachment regions of genes to the nuclear matrix. J.Cell.Biochem. 52:14-22.

27. Bradbury, E.M. (1992) Reversible histone modifications and the chromosome cell cycle. BioEssays 14:9-16.

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

29. Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E. and Arnold, H.H. (1989) A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. EMBO J. 8:701-709.

30. Braunstein, M., Rose, A.B., Holmes, S.G., Allis, C.D. and Broach, J.R. (1993) Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. Genes Dev. 7:592-604.

31. Brent, R. and Ptashne, M. (1985) A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell 43:729-736.

32. Brown, C.J., Lafreniere, R.G., Powers, V.E., Sebastio, G., Ballabio, A., Pettigrew, A.L., Ledbetter, D.H., Levy, E., Craig, I.W. and Willard, H.F. (1991) Localization of the X inactivation centre on the human X chromosome in Xq13. nature 349:82-84.

33. Brown, C.J. and Willard, H.F. (1990) Localization of a gene that escapes inactivation to the X chromosome proximal short arm:implications for X inactivation. Am.J.Hum.Genet. 46:273-279.

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

35. Burtis, K.C., Thummel, C.S., Jones, C.W., Karim, F.D. and Hogness, D.S. (1990) The Drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene that encodes two ets-related proteins. Cell 61:85-99.

36. Busch, H. (1984) Ubiquitination of proteins. Methods Enzymol. 106:238-262.

37. Busch, S.J. and Sassone-Corsi, P. (1990) Dimers, leucine zippers and DNA-binding domains. Trends Genet. 6:36-40.

38. Carey, M., Lin, Y.-S., Green, M.R. and Ptashne, M. (1990) A mechanism for synergistic activation of a mammalian gene by Gal4 derivatives. nature 345:361-364.

39. Carr, C.S. and Sharp, P.A. (1990) A helix-loop-helix protein related to the immunoglobulin E box-binding proteins. Mol.Cell.Biol. 10:4384-4388.

40. Carthew, R.W., Chodosh, L.A. and Sharp, P.A. (1985) An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell 43:439-448.

41. Caudy, M., Grell, E.H., Damblych, C., Ghysen, A., Jan, L.Y. and Jan, Y.N. (1988) The maternal sex determination gene daughterless has zygotic activity necessary for the formation of peripheral neurons in Drosophila. Genes Dev. 2:843-852.

42. Chambers, A., Stanway, C., Kingsman, A.J. and Kingsman, S.M. (1988) The UAS of the yeast PGK gene is composed of multiple functional elements. Nucl.AcidsRes. 16:8245-8260.

43. Chambers, A., Tsang, J.S.H., Stanway, C., Kingsman, A.J. and Kingsman, S.M. (1989) Trancriptional control of the Saccharomyces cerevisiae PGK gene by RAP1. Mol.Cell.Biol. 9:5516-5524.

44. Chen, C. and Okayama, H. (1987) High-efficiency transformation of mammalian cells by plasmid DNA. Mol.Cell.Biol. 7:2745-2752.

45. Chiang, C.-M., Ge, H., Wang, Z., Hoffman, A. and Roeder, R.G. (1993) Unique TATA-binding protein-containing complexes and cofactors involved in transcription by RNA polymerases II and III. EMBO J. 12:2749-2762.

46. Chodosh, L.A., Baldwin, A.S., Carthew, R.W. and Sharp, P.A. (1988) Human CCAAT-binding proteins have heterologous subunits. Proc.Natl.Acad.Sci.USA 53:11-24.

47. Chodosh, L.A., Carthew, R.W. and Sharp, P.A. (1986) A single polypeptide possesses the binding and activities of the adenovirus major late transcription factor. Mol.Cell.Biol. 6:4723-4733.

48. Chu, C.C., Templin, A. and Clark, A.J. (1989) Suppression of a frameshift mutation in the recE gene of Escherichia coli K-12 occurs by gene fusion. J.Bacteriol. 171:2101-2109.

49. Church, G.M., Ephrussi, A., Gilbert, W. and Tonegawa, S. (1985) Cell-type-specific contacts to immunoglobulin enhancers in nuclei. nature 313:798-801.

50. Cockerill, P.N. and Garrard, W.T. (1986) Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell 44:272-282.

51. Colgan, J., Wampler, S. and Manley, J.L. (1993) Interaction between a transcriptional activator and transcription factor IIB in vivo. nature 362:549-553.

52. Conaway, R.C., Garrett, K.P., Hanley, J.P. and Conaway, J.W. (1991) Mechanism of promoter selection by RNA pol.II:mammalian transcription factors α and $\beta\delta$ promote entry of polymerase into the preinitiation complex. Proc.Natl.Acad.Sci.USA 88:6205-6209.

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

54. Cronmiller, C., Schedl, P. and Cline, T.W. (1988) Molecular characterization of daughterless, a Drosophila sex determination gene with multiple roles in development. Genes Dev. 2:1666-1676.

55. Crowe, J.S., Cooper, H.J., Smith, M.A., Sims, M.J., Parker, D. and Gewert, D. (1991) Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinase K digestion. Nucl.AcidsRes. 19:184.

56. Crowley, T.E., Hoey, T., Liu, J.-K., Jan, Y.-N., Jan, L.-Y. and Tjian, R. (1993) A new factor related to TATA-binding protein has highly restricted expression patterns in Drosophila. nature 361:557-561.

57. Curran, T. and Franza, B.R.,Jr. (1988) Fos and Jun:the AP-1 connection. Cell 55:395-397.

58. Dalton, S. and Treisman, R. (1992) Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element. Cell 68:597-612.

59. Davis, R.L., Cheng, P.F., Lassar, A.B. and Weintraub, H. (1990) The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell 60:773-746.

60. Davis, R.L., Weintraub, H. and Lassar, A.B. (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 51:987-1000.

61. De Belle, I., Walker, P.R., Smith, I.C.P. and Sikorska, M. (1991) Identification of a multiprotein complex interacting with the c-fos serum response element. Mol.Cell.Biol. 11:2752-2759.

62. Desser, G.N. *Cellular and Molecular Biology of Intermediate Filaments*, New York:Plenum Press, 1990. pp. 129-145.

63. Dignam, J.D., Martin, P.L., Shastry, B.S. and Roeder, R.G. Eukaryotic gene transcription with purified components. In: *Methods in Enzymology*, Academic Press, Inc., 1983, p. 582-598.

64. Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987) A multiplicity of CCAAT box-binding proteins. Cell 50:863-872.

65. Dostatni, N., Lambert, P.F., Sousa, R., Ham, J., Howley, P.M. and Yaniv, M. (1991) The functional BPV-1 E2 trans-activation protein can act as a repressor by preventing formation of the initiation complex. Genes Dev. 5:1657-1671.

66. Drolet, D.W., Scully, K.M., Simmons, D.M., Wegner, M., Chu, K., Swanson, L.W. and Rosenfeld, M.G. (1991) TEF, a transcription factor expressed specifically in the anterior pituitary during embryogenesis, defines a new class of leucine zipper proteins. Genes Dev. 5:1739-1753.

67. Dudek, H., Tantravati, R.V., Rao, V.N., Reddy, E.S. and Reddy, E.P. (1992) Myb and Ets proteins cooperate in transcriptional activation of the mim-1 promoter. Proc.Natl.Acad.Sci.USA 89:1291-1295.

68. Dworetzky, S.I., Wright, K.L., Fey, E.G., Penman, S., Lian, J.B., Stein, J.L. and Stein, G.S. (1992) Sequence-specific DNA-binding proteins are components of a nuclear matrix-attachment site. Proc.Natl.Acad.Sci.USA 89:4178-4182.

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

70. Dynlacht, B.D., Hoey, T. and Tjian, R. (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell 66:563-576.

71. Edmondson, D.G. and Olson, E.N. (1989) A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Genes Dev. 3:628-640.

72. Ellis, H., Spann, D.R. and Posakony, J.W. (1990) Extramacrochaetae, a negative regulator of sensory organ development in Drosophila, defines a new class of helix-loop-helix proteins. Cell 61:27-38.

73. Endo, S., Takagi, N. and Sasaki, M. (1982) The late-replicating X chromosome in digynous mouse triploid embryos. Dev.Genet. 3:165-176.

74. Felsenfeld, G. (1992) Chromatin as an essential part of the transcriptional mechanism. nature 355:219-224.

75. Fisher, D.E., Parent, L.A. and Sharp, P.A. (1992) Myc/Max and other helix-loop-helix/leucine zipper proteins bend DNA toward the minor groove. Proc.Natl.Acad.Sci.USA 89:11779-11783.

76. Flanagan, P.M., Kelleher, R.J.,III, Sayre, M.H., Tschochner, H. and Kornberg, R.D. (1991) A mediator required for activation of RNA polymerase II transcription in vitro. nature 350:436-438.

77. Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z.F. and Reinberg, D. (1991) The small subunit of transcription factor IIF recruits RNA polymerase into the preinitiation complex. Proc.Natl.Acad.Sci.USA 88:9999-10003.

78. Foisner, R. and Gerace, L. (1993) Integral membrane proteins of the nuclear envelope interact with lamins and chromosomes, and binding is modulated by mitotic phosphorylation. Cell 73:1267-1279.

79. Foulkes, N.S., Borrelli, E. and Sassone-Corsi, P. (1991) CREM gene: use of alternative DNA-binding domains generates multiple antagonists of cAMP-induced transcription. Cell 64:739-749.

80. Frampton, J., Leutz, A., Gibson, T.J. and Graf, T. (1991) DNA-binding domain ancestry. nature 342:134.

81. Fried, M. and Crothers, D.M. (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucl.AcidsRes. 9:6505-6523.

82. Fujiwara, S., Koizumi, S., Fisher, R.J., Bhat, N.K. and Papas, J.S. (1990) Phosphorylation of the ETS-2 protein:regulation by T-cell antigen receptor-CD3 complex. Mol.Cell.Biol. 3:1249-1253.

83. Garner, M.M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions:application to components of the Escherichia coli lactose operon regulatory system. Nucl.AcidsRes. 9:3047-3059.

84. Garrell, J. and Modolell, J. (1990) The Drosophila extramacrochaetae locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. Cell 61:39-48.

85. Gartenberg, M.R., Ampe, C., Steitz, T.A. and Crothers, D.M. (1990) Molecular characterization of the GCN4-DNA complex. Proc.Natl.Acad.Sci.USA 87:6034-6038.

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

87. Gasch, A., Hoffman, A., Horikoshi, M., Roeder, R.G. and Chau, N.-H. (1990) Arabidopsis thaliana contains two genes for TFIID. nature 346:390-394.

88. Gasser, S.M. and Laemmli, U.K. (1986) Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of D.melanogaster. Cell 46:521-530.

89. Gebara, M.M. and McCarrey, J.R. (1992) Protein-DNA interactions associated with the onset of testis-specific expression of the mammalian Pgk-2 gene. Mol.Cell.Biol. 12:1422-1431.

90. Gerace, L. and Burke, B. (1988) Functional organization of the nuclear envelope. Ann.Rev.Cell Biol. 4:335-374.

91. Gerster, T. and Roeder, R.G. (1988) A herpes virus transactivating protein interacts with transcription factor OTF-1 and other cellular proteins. Proc.Natl.Acad.Sci.USA 85:6347-6351.

92. Getzenberg, R.H., Pienta, K.J., Ward, W.S. and Coffey, D.S. (1991) Nuclear structure and the three-dimensional organization of DNA. J.Cell.Biochem. 47:289-299.

93. Giese, K., Amsterdam, A. and Grosschedl, R. (1991) DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. Genes Dev. 5:2567-2578.

94. Giese, K., Cox, J. and Grosschedl, R. (1992) The HMG domain of lymphoid enhancer factor I bends DNA and facilitates assembly of functional nucleoprotein structures. Cell 69:185-195.

95. Gil, G., Smith, J.R., Goldstein, J.L., Slaughter, C.A., Orth, K., Brown, M.S. and Osborne, T.F. (1988) Multiple genes encode nuclear factor 1-like proteins that bind to the promoter for 3-hyroxy-3-methylglutaryl-coenzyme A reductase. Proc.Natl.Acad.Sci.USA 85:8963-8967.

96. Gill, G. and Ptashne, M. (1987) Mutants of GAL4 protein altered in an activation function. Cell 51:121-126.

97. Goldman, M.A. (1988) The chromatin domain as a unit of gene regulation. BioEssays 9:50-55.

98. Goldman, M.A., Stokes, K.R., Idzerda, R.L., McKnight, G.S., Hammer, R.E., Brinster, R.L. and Gartler, S.M. (1987) A chicken transferrin gene in transgenic mice escapes X-chromosome inactivation. Science 236:593-595.

99. Goto, M., Masamune, Y. and Nakanishi, Y. (1993) A factor stimulating transcription of the testis-specific Pgk-2 gene recognizes a sequence similar to the binding site for a transcription inhibitor of the somatic-type Pgk-1 gene. Nucl.AcidsRes. 21:209-214.

100. Goto, M., Tamura, T., Mikoshiba, K., Masamune, Y. and Nakanishi, Y. (1991) Transcription inhibition of the somatic-type phosphoglycerate kinase 1 gene in vitro by a testis-specific factor that recognizes a sequence similar to the binding site for Ets oncoproteins. Nucl.AcidsRes. 19:3959-3963.

101. Gounari, F., De Francesco, R., Schmitt, J., van der Vliet, P.C., Cortese, R. and Stunnenberg, H. (1990) Amino-terminal domain of NF1 binds to DNA as a dimer and activates adenovirus DNA replication. EMBO J. 9:559-566.

102. Goyal, N., Knox, J. and Gronostajski, R.M. (1990) Analysis of multiple forms of nuclear factor 1 in human and murine cell lines. Mol.Cell.Biol. 10:1041-1048.

103. Grant, S.G. and Chapman, V.M. (1988) Mechanisms of X-chromosome regulation. Annu.Rev.Genet. 22:199-233.

104. Gregor, P.D., Sawadogo, M. and Roeder, R.G. (1990) The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. Genes Dev. 4:1730-1740.

105. Gronostajski, R.M. (1987) Site-specific DNA binding of nuclear factor 1:effect of the spacer region. Nucl.AcidsRes. 15:5545-5559.

106. Gross, D.S. and Garrard, W.T. (1988) Nuclease hypersensitive sites in chromatin. Annu.Rev.Biochem. 57:159-197.

107. Gruda, M.C., Zabolotny, J.M., Xiao, J.H., Davidson, I. and Alwine, J.C. (1993) Transcriptional activation by simian virus 40 large T antigen:interactions with multiple components of the transcription complex. Mol.Cell.Biol. 13:961-969.

108. Guarente, L. (1988) UASs and enhancers:common mechanism of transcriptional activation in yeast and mammals. Cell 52:303-305.

109. Gutman, A. and Wasylyk, B. (1993) Nuclear targets for transcription regulation by oncogenes. Trends Genet. 7:49-54.

110. Hagemeier, C., Bannister, A.J., Cook, A. and Kouzardes, T. (1993) The activation domain of transcription factor PU.1 binds the retinoblastoma (RB) protein and the transcription factor TFIID in vitro:RB shows sequence similarity to TFIID and TFIIB. Proc.Natl.Acad.Sci.USA 90:1580-1584.

111. Hai, T., Liu, F., Coukos, W.J. and Green, M.R. (1989) Transcription factor ATF cDNA clones:an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. Genes Dev. 3:2083-2090.

112. Halazonetis, T.D. and Kandil, A.N. (1991) Determination of the c-MYC DNA-binding site. Proc.Natl.Acad.Sci.USA 88:6162-6166.

113. Harrison, S.C. (1991) A structural taxonomy of DNA-binding domains. nature 353:715-719.

114. Hatamochi, A., Golumbek, P.T., van Schaftingen, E. and de Crombrugghe, B. (1988) A CCAAT DNA binding factor consisting of two different components that are both required for DNA binding. J.Biol.Chem. 263:5940-5947.

115. Hateboer, G., Timmers, H.T.M., Rustgi, A.K., Billaud, M., van 't Veer, L. and Bernards, R. (1993) TATA-binding protein and the retinoblastoma gene product bind to overlapping epitopes on c-Myc and adenovirus E1A protein. Proc.Natl.Acad.Sci.USA 90:8489-8493.

116. He, X. and Rosenfeld, M.G. (1991) Mechanisms of complex transcriptional regulation:implications for brain development. Neuron 7:183-196.

117. Hebbes, T.R., Thorne, A.W. and Crane-Robinson, C. (1988) A direct link between core histone acetylation and transcriptionally active chromatin. EMBO J. 7:1395-1402.

118. Hill, C.S., Parkman, L.C. and Thomas, J.O. (1990) Phosphorylation at clustered -Ser-Pro-X-Lys/Arg- motifs in sperm-specific histones H1 and H2B. EMBO J. 9:805-813.

119. Hisatake, K., Hasegawa, S., Takada, R., Nakatani, Y., Horikoshi, M. and Roeder, R.G. (1993) The p250 subunit of native TATA box-binding factor TFIID is the cell-cycle regulatory protein CCG1. nature 362:179-181.

120. Hoey, T., Weinzierl, R.O.J., Gill, G., Chen, J.-L., Dynlacht, B.D. and Tjian, R. (1993) Molecular cloning and functional analysis of Drosophila TAF110 reveal properties expected of coactivators. Cell 72:247-260.

121. Hoffman, A., Sinn, E., Yamamato, T., Wang, J., Roy, A., Horikoshi, M. and Roeder, R.G. (1990) Highly conserved core domain and unique N terminus with presumptive regulatory motifs in a human TATA factor (TFIID). nature 346:387-390.

122. Horikoshi, M., Bertuccioli, C., Takada, R., Wang, J., Yamamoto, T. and Roeder, R.G. (1992) Transcription factor TFIID induces DNA bending upon binding to the TATA element. Proc.Natl.Acad.Sci.USA 89:1060-1064.

123. Horikoshi, M., Hai, T., Lin, Y.-S., Green, M.R. and Roeder, R.G. (1988) Transcription factor ATF interacts with the TATA factor to facilitate establishment of a preinitiation complex. Cell 54:1033-1042.

124. Hornstra, I.K. and Yang, T.P. (1992) Multiple in vivo footprints are specific to the active allele of the X-linked human hypoxanthine phosphoribosyltransferase gene 5'-region:implications for X chromosome inactivation. Mol.Cell.Biol. 12:5345-5354.

125. Hu, Y.F., Lüscher, B., Admon, A. and Tjian, R. (1990) Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. Genes Dev. 4:1741-1752.

126. Huang, H.C. and Cole, R.D. (1984) The distribution of H1 histone is non-uniform in chromatin and correlates with different degrees of condensation. J.Biol.Chem. 259:14237-14242.

127. Hutcheon, T., Dixon, G.H. and Levy Wilson, B. (1980) Transcriptionally active mononucleosomes from trout testis are heterogeneous in composition. J.Biol.Chem. 255:681-685.

128. Huynh, T.V., Young, R.A. and Davis, R.W. *DNA-cloning-A Practical Approach*, Oxford:IRL Press, 1985. pp. 49-78.

129. Inostroza, J., Flores, O. and Reinberg, D. (1991) Factors involved in specific transcription by mammalian RNA polymerase II. J.Biol.Chem. 266:9304-9308.

130. Inostroza, J.A., Merimelstein, F.H., Ha, I., Lane, W.S. and Reinberg, D. (1992) Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription. Cell 70:477-489.

131. Ish-Horowicz, D. and Burke, J.F. (1981) Rapid and efficient cosmid cloning. Nucl.AcidsRes. 9:2989-2998.

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

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

134. Jones, N. (1990) Transcriptional regulation by dimerization:two sides of an incestuous relationship. Cell 61:9-11.

135. Jones-Villeneuve, E.M.V., McBurney, M.W., Rogers, K.A. and Kalnins, V.I.J. (1982) Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. J.Cell Biol. 94:253-262.

136. Kadonaga, J.T., Carner, K.R., Masiarz, F. and Tjian, R. (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 51:1079-1090.

137. Kadonaga, J.T., Courey, A.J., Ladika, J. and Tjian, R. (1988) Distinct regions of Sp1 modulate DNA binding and transcriptional activation. Science 242:1566-1569.

138. Kageyama, R., Merlino, G.T. and Pastan, I. (1989) Nuclear factor ETF specifically stimulates transcription from promoters without a TATA box. J.Biol.Chem. 264:15508-15514.

139. Kamel, R.A., Berg, K., Schwarzfischer, F. and Wischerath, H. (1975) First determination of the isozyme patterns of phosphoglycerate mutases (E.C.2.7.5.3) and phosphoglycerate kinases (E.C.2.7.2.3) in human tissues. Humangenetik 27:53-55.

140. Karim, F.D., Urness, L.D., Thummel, C.S., Klemsz, M.J., McKercher, S.R., Celada, A., van Beveren, C., Maki, R.A., Gunther, C.V., Nye, J.A. and Graves, B.J. (1990) The ETS-domain:a new DNA-binding motif that recognizes a purine-rich core DNA sequence. Genes Dev. 4:1451-1453.

141. Karpov, V.L., Preobrazhenskaya, V. and Mirzabekov, A.P. (1984) Chromatin structure of hsp70 genes, activated by heat shock:selective removal of histones from the coding region and their absence from the 5' region. Cell 36:423-431.

142. Kelleher, R.J.,III, Flanagan, P.M. and Kornberg, R.D. (1990) A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. Cell 61:1209-1215.

143. Kelly, M., Burke, J., Smith, M., Klar, A. and Beach, D. (1988) Four mating-type genes control sexual differentiation in the fission yeast. EMBO J. 7:1537-1547.

144. Kerkhoff, E., Bister, K. and Klempnauer, K-H. (1991) Sequence-specific DNA binding by myc proteins. Proc.Natl.Acad.Sci.USA 88:4323-4327.

145. Kerppola, T.K. and Curran, T. (1991) Fos-Jun heterodimers and Jun homodimers bend DNA in opposite orientations:implications for transcription factor cooperativity. Cell 66:317-326.

146. Kerppola, T.K. and Curran, T. (1993) Selective DNA bending by a variety of bZIP proteins. Mol.Cell.Biol. 13:5479-5489.

147. Kerr, L.D., Ransone, L.J., Wamsley, P., Schmitt, M.J., Boyer, T.G., Zhou, Q., Berk, A.J. and Verma, I.M. (1993) Association between proto-oncoprotein Rel and TATA-binding protein mediates transcriptional activation by Nf-kappaB. nature 365:412-419.

148. King, I.N., de Soyza, T., Catanzaro, D.F. and Lavin, T.N. (1993) Thyroid hormone receptor-induced bending of specific DNA sequences is modified by an accessory factor. J.Biol.Chem. 268:495-501.

149. Klehr, D., Maass, K. and Bode, J. (1991) Scaffold-attached regions from the human interferon β domain can be used to enhance the stable expression of genes under the control of various promoters. Biochem 30:1264-1270.

150. Klemsz, M.J., McKerder, S.R., Celada, A., van Beveren, C. and Maki, R.A. (1990) The macrophage and B cell specific transcription factor PU.1 is related to the ets oncogene. Cell 61:113-124.

151. Kola, I., Brookes, S., Green, A.R., Garber, R., Tymms, M., Papas, T.S. and Seth, A. (1993) The Ets1 transcription factor is widely expressed during murine embryo development and is associated with mesodermal cells involved in morphogenetic processes such as organ formation. Proc.Natl.Acad.Sci.USA 90:7588-7592.

152. Kouzarides, T. and Ziff, E. (1988) The role of the leucine zipper in the fos-jun interaction. nature 336:646-651.

153. Kouzarides, T. and Ziff, E. (1989) Leucine zippers of fos, jun and GCN4 dictate dimerization specificity and thereby control DNA binding. nature 340:568-571.

154. Kozak, L.P., McLean, G.K. and Eicher, E.M. (1974) X linkage of phosphoglycerate kinase in the mouse. Biochem.Genet. 11:41-47.

155. Krumlauf, R., Chapman, V.M., Hammer, R.E., Brinster, R. and Tilghman, S.M. (1986) Differential expression of alpha-fetoprotein genes on the inactive X chromosome in extraembryonic and somatic tissues of a transgenic mouse line. nature 319:224-226.

156. Laemmli, U.K., Käs, E., Poljak, L. and Adachi, Y. (1992) Scaffold-associated regions:cis-acting determinants of chromatin structural loops and functional domains. Current Opinions in Genetics and Development 2:275-285.

157. LaMarco, K., Thompson, C.C., Byers, B.P., Walton, E.M. and McKnight, S.L. (1991) Identification of ets- and notch- related subunits in GA binding protein. Science 253:789-792.

158. Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) The leucine zipper:a hypothetical structure common to a new class of DNA binding proteins. Science 240:1759-1764.

159. Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D. and Weintraub, H. (1991) Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. Cell 66:305-315.

160. Laybourn, P.J. and Kadonaga, J.T. (1991) Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. Science 254:238-245.

161. Lee, D.Y., Hayes, J.J., Pruss, D. and Wolffe, A.P. (1993) A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell 72:73-84.

162. Lee, J.-S., Galvin, K.M. and Shi, Y. (1993) Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. Proc.Natl.Acad.Sci.USA 90:6145-6149.

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

164. Levine, M. and Manley, J.L. (1989) Transcriptional repression of eukaryotic promoters. Cell 59:405-408.

165. Lewin, B. (1990) Commitment and activation at Pol II promoters:a tail of protein-protein interaction. Cell 61:1161-1164.

166. Li, R., Knight, J.D., Jackson, S.P., Tjian, R. and Botchan, M.R. (1991) Direct interaction between Sp1 and the BPV enhancer E2 protein mediates synergistic activation of transcription. Cell 65:493-505.

167. Lieberman, P.M. and Berk, A.J. (1991) The Zta trans-activator protein stabilizes TFIID association with promoter DNA by direct protein:protein interaction. Genes Dev. 5:2441-2454.

168. Lilley, D.M.J. (1992) HMG has DNA wrapped up. nature 357:282-283.

169. Lin, Y.-S. and Green, M.R. (1991) Mechanism of action of an acidic transcriptional activator in vitro. Cell 64:971-981.

170. Liu, X., Miller, C.W., Koeffler, P.H. and Berk, A.J. (1993) The p53 activation domain binds the TATA box-binding polypeptide in Holo-TFIID, and a neighboring p53 domain inhibits transcription. Mol.Cell.Biol. 13:3291-3300.

171. Lock, L.F., Takagi, N. and Martin, G.R. (1987) Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell 48:36-46.

172. Lu, H., Zawel, L., Fisher, L., Egly, J.-M. and Reinberg, D. (1992) Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. nature 358:641-645.

173. Lu, X.-P., Eberhardt, N.L. and Pfahl, M. (1993) DNA bending by retinoid X receptor-containing retinoid and thyroid hormone receptor complexes. Mol.Cell.Biol. 13:6509-6519.

174. Luderus, M.E.E., deGraaf, A., Mattia, E., den Blaauwen, J.L., Grande, M.A., de Jong, L. and van Driel, R. (1992) Binding of matrix attachment regions to lamin B1. Cell 70:949-959.

175. Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R. and Sigler, P.B. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. nature 352:497-505.

176. Lyon, M.F. (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). nature 190:372-373.

177. Lyon, M.F. (1992) Some milestones in the history of X-chromosome inactivation. Annu.Rev.Genet. 26:17-28.

178. Mack, D.H., Vartikar, J., Pipas, J.M. and Laimins, L.A. (1993) Specific repression of TATA-mediated but not initiator-mediated transcription by wild-type p53. nature 363:281-283.

179. Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987) Regulation of inducible and tissue-specific gene expression. Science 236:1237-1245.

180. Marko, M.A., Chipperfield, R. and Birnboim, H.C. (1982) A procedure for the large scale isolation of highly purified plasmid DNA using alkaline extraction and binding to glass powder. Anal.Biochem. 121:382-387.

181. Martin, M.E., Yang, X.Y. and Folk, W.R. (1992) Expression of a 91 kD PEA3-binding protein is downregulated during differentiation of F9 embryonal carcinoma cells. Mol.Cell.Biol. 12:2213-2221.

182. Maxam, A.M. and Gilbert, W. Sequencing end-labelled DNA with base-specific chemical cleavages. In: *Methods in Enzymology*, Academic Press,Inc., 1980, p. 499-560.

183. McBurney, M.W. (1988) X chromosome inactivation:a hypothesis. BioEssays 9:85-88.

184. McBurney, M.W., Jones-Villeneuve, E.M.V., Edwards, M. and Anderson, P. (1982) Controlled differentiation of teratocarcinoma cells in culture. nature 299:165-167.

185. McBurney, M.W. and Rogers, B.J. (1982) Isolation of male murine embryonal carcinoma cells and their chromosome replication patterns. Dev.Biol. 89:503-508.

186. McBurney, M.W., Staines, W.A., Boekelheide, K., Parry, D., Jardine, K. and Pickavance, L. (1993) Autosomal transgene driven by the promoter of an X-linked gene (Pgk-1) is expressed widely and neither inactivated nor imprinted. submitted

187. McBurney, M.W., Sutherland, L.C., Adra, C.N., Leclair, B., Rudnicki, M.A. and Jardine, K. (1991) The mouse Pgk-1 gene promoter contains an upstream activator sequence. Nucl.AcidsRes. 19:5755-5761.

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

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

190. McQuillan, J.J., Rosen, G.D., Birkenmeier, T.M. and Dean, D.C. (1991) Identification of a protein that interacts with the nuclear factor-1 (NF-1) binding site in cells that do not express NF-1:comparison to NF-1, cellular distribution, and effect on transcription. Nucl.AcidsRes. 19:6627-6631.

191. Means, A.L. and Farnham, P.J. (1990) Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. Mol.Cell.Biol. 10:653-661.

192. Megee, P.C., Morgan, B.A., Mittman, B.A. and Smith, M.M. (1990) Genetic analysis of histone H4:essential role of lysines subject to reversible acetylation. Science 247:841-845.

193. Meisterernst, M. and Roeder, R.G. (1991) Family of proteins that interact with TFIID and regulate promoter activity. Cell 67:557-567.

194. Meisterernst, M., Roy, A.L., Lieu, H.M. and Roeder, R.G. (1991) Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. Cell 66:981-993.

195. Merino, A., Madden, K.R., Lane, W.S., Champoux, J.J. and Reinberg, D. (1993) DNA topoisomerase I is involved in both repression and activation of transcription. nature 365:227-232.

196. Mermod, N., O'Neill, E.A., Kelly, T.J. and Tjian, R. (1989) The proline-rich transcriptional activation of CTF/NF-1 is distinct from the replication and DNA binding domain. Cell 58:741-753.

197. Mirkovitch, J., Mirault, M.-E. and Laemmli, U.K. (1984) Organization of the higher-order chromatin loop:specific DNA attachment sites on nuclear scaffold. Cell 39:223-232.

198. Mitchell, P. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245:371-378.

199. Mitchell, P.J. and Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245:371-378.

200. Mohandas, T., Geller, R.L., Yen, P.H., Rosendorf, J., Bernstein, R., Yoshida, A. and Shapiro, L.J. (1987) Cytogenetic and molecular studies on a recombinant human X chromosome:implications for the spread of X chromosome inactivation. Proc.Natl.Acad.Sci.USA 84:4954-4958.

201. Monk, M. and Harper, M.I. (1979) Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. nature 281:311-313.

202. Murphy, S., Yoon, J.-B., Gerster, T. and Roeder, R.G. (1992) Oct-1 and Oct-2 potentiate functional interaction of a transcription factor with the proximal sequence element of snRNA genes. Mol.Cell.Biol. 12:3247-3261.

203. Murre, C., McCaw, P.S. and Baltimore, D. (1989a) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell 56:777-783.

204. Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lassar, A.B., Weintraub, H. and Baltimore, D. (1989b) Interactions between heterologous helix-loop-helix proteins generates complexes that bind specifically to a common DNA sequence. Cell 58:537-544.

205. Müller, M.M., Gerster, T. and Schaffner, W. (1988) Enhancer sequences and the regulation of gene transcription. Eur.J.Biochem. 176:485-495.

206. Nakatani, Y. (1990) A downstream initiation element required for efficient TATA box binding and in vitro function of TFIID. nature 348:86-88.

207. Nakshatri, H., Pater, M.M. and Pater, A. (1990) Activity and enhancer binding-factors for JC virus regulatory elements in differentiating embryonal carcinoma cells. Virology 177:784-789.

208. Nardulli, A.M. and Shapiro, D.J. (1992) Binding of the estrogen receptor DNA-binding domain to the estrogen response element induces DNA bending. Mol.Cell.Biol. 12:2037-2042.

209. Neuberg, M., Adamkiewicz, J., Hunter, J.B. and Muller, R. (1989) A fos protein containing the jun leucine zipper forms a homodimer which binds to the AP-1 binding site. nature 341:243-245.

210. Nilsson, P., Hallberg, B., Thornell, A. and Grundström, T. (1989) Mutant analysis of protein interactions with a nuclear factor 1 binding site in the SL3-3 virus enhancer. Nucl.AcidsRes. 17:4061-4075.

211. Norton, P.A. and Coffin, J.M. (1985) Bacterial B-gal as a marker of Rous Sarcoma virus gene expression and replication. Mol.Cell.Biol. 5:281-290.

212. Nye, J.A., Petersen, J.M., Gunther, C.V., Jonsen, M.D. and Graves, B.J. (1992) Interaction of murine Ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. Genes Dev. 6:975-990.

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

214. Ohkuma, Y., Horikoshi, M., Roeder, R.G. and Desplan, C. (1990) Engrailed, a homeodomain protein, can repress in vitro transcription by competition with the TATA box-binding protein transcription factor IID. Proc.Natl.Acad.Sci.USA 89:2289-2293.

215. Oikarinen, J. and Mannermaa, R.-M. (1990) The predicted molecular structure suggests that CTF/ NF-1 may function as a histone acetylase. FEBS Lett, 273:11-14.

216. Oliva, R., Bazett-Jones, D.P., Locklear, L. and Dixon, G.H. (1990) Histone hyperacetylation can induce unfolding of the nucleosome core particle. Nucl.AcidsRes. 18:2739-2747.

217. Pascal, E. and Tjian, R. (1991) Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. Genes Dev. 5:1646-1656.

218. Pegoraro, B. and Lee, C-Y. (1978) Purification and characterization of two isozymes of 3-phosphoglycerate kinase from the mouse. Biochim.Biophys.Acta 52:423-433.

219. Peterson, M.G., Inostroza, J., Maxon, M.E., Flores, O., Admon, A., Reinberg, D. and Tjian, R. (1991) Structure and functional properties of human general transcription factor IIE. nature 354:369-373.

220. Peterson, M.G., Tanese, N., Pugh, B.F. and Tjian, R. (1990) Functional domains and upstream activation properties of cloned human TATA binding protein. Science 248:1625-1630.

221. Peterson, M.G. and Tjian, R. (1992) The tell-tail trigger. nature 358:620-621.

222. Pfeifer, K., Kim, K.-S., Kogan, S. and Guarente, L. (1989) Functional dissection and sequence of yeast HAP1 activator. Cell 56:291-301.

223. Pfeiffer, G.P. and Riggs, A.D. (1991) Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permeabilized cells using DNase 1 and ligation-mediated PCR. Genes Dev. 5:1102-1113.

224. Pfeiffer, G.P., Tanguay, R.L., Steigerwald, S.D. and Riggs, A.D. (1990) In vivo footprint and methylation analysis by PCR-aided genomic sequencing:comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. Genes Dev. 4:1277-1287.

225. Phi-Van, L., von Kries, J.P., Ostertag, W. and Strätling, W.H. (1990) The chicken lysozyme 5' matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. Mol.Cell.Biol. 10:2302-2307.

226. Pienta, K.J., Getzenberg, R.H. and Coffey, D.S *CRC reviews on eukaryotic gene expression*, 1991. pp. 355-385.

227. Pognonec, P., Boulukos, K.E., Bosselut, R., Boyer, C., Schmitt-Verhulst, A.M. and Ghysdael, J. (1990) Identification of a Ets1 variant protein unaffected in its chromatin and in vitro DNA binding capacities by T cell antigen receptor triggering and intracellular calcium rises. Oncogene 5:603-610.

228. Ptashne, M. (1988) How eukaryotic transcriptional activators work. nature 335:683-689.

229. Ptashne, M. and Gann, A.A.F. (1990) Activators and targets. nature 346:329-331.

230. Ptashne, M., Jeffrey, A., Johnson, A.D., Maurer, R., Meyer, B.J., Pabo, C.O., Roberts, T.M. and Sauer, R.T. (1980) How the lambda repressor and Cro work. Cell 19:1-11.

231. Pugh, B.F. and Tjian, R. (1990) Mechanism of transcriptional activation by Sp1:evidence for coactivators. Cell 61:1187-1197.

232. Pugh, B.F. and Tjian, R. (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev. 5:1935-1945.

233. Pugh, B.F. and Tjian, R. (1992) Diverse transcriptional functions of the multisubunit eukaryotic TFIID complex. J.Biol.Chem. 267:679-682.

234. Rao, V.N., Papas, T.S. and Reddy, E.S.P. (1987) erg, a human ets-related gene on chromosome 21:alternative splicing, polyadenylation, and translation. Science 237:635-639.

235. Rastan, S. (1983) Non-random X-chromosome inactivation in mouse X-autosome translocation embryos - location of the inactivation center. J.Embryol.Exp.Morphol. 78:1-22.

236. Rastan, S. and Robertson, E.J. (1985) X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. J.Embryol.Exp.Morphol. 90:279-388.

237. Reddy, E.S.P., Rao, V.N. and Papas, T.S. (1987) The erg gene:a human gene related to the ets oncogene. Proc.Natl.Acad.Sci.USA 84:6131-6135.

238. Reddy, E.S.P. and Rao, V.N. (1991) erg, an ets-related gene, codes for sequence-specific transcriptional activators. Oncogene 6:2285-2289.

239. Reeves, R. (1984) Transcriptionally active chromatin. Biochim.Biophys.Acta 782:343-393.

240. Reik, A., Schütz, G. and Stewart, A.F. (1991) Glucocorticoids are required for establishment and maintenance of an alteration in chromatin structure:induction leads to a reversible disruption of nucleosomes over an enhancer. EMBO J. 10:2569-2576.

241. Rigaud, G., Roux, J., Pictet, R. and Grange, T. (1991) In vivo footprinting of Rat TAT gene:dynamic interplay between the glucocorticoid receptor and a liver-specific factor. Cell 67:977-986.

242. Ristiniemi, J. and Oikarinen, J. (1989) Histone H1 binds to the putative nuclear factor 1 recognition sequence in the mouse α2(I) collagen promoter. J.Biol.Chem. 264:2164-2174.

243. Robinson, M.O., McCarrey, J.R. and Simon, M.I. (1989) Transcriptional regulatory regions of testis-specific PGK2 defined in transgenic mice. Proc.Natl.Acad.Sci.USA 86:8437-8441.

244. Robson, R.M. (1989) Intermediate filaments. Curr.Opin.Cell Biol. 1:36-43.

245. Rojo, F. and Salas, M. (1991) A DNA curvature can substitute phage ϕ29 regulatory protein p4 when acting as a transcriptional repressor. EMBO J. 10:3429-3438.

246. Roman, C., Cohen, L. and Calame, K. (1991) A dominant negative form of transcription activator mTFE3 created by differential splicing. Science 254:94-97.

247. Romig, H., Fackelmayer, F.O., Renz, A., Ramsperger, U. and Richter, A. (1992) Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. EMBO J. 11:3431-3440.

248. Roselli-Rehfuss, L., Ye, F., Lissemor, J.L. and Sullivan, D.T. (1992) Structure and expression of the phosphoglycerate kinase (PGK) gene of Drosophila melanogaster. Mol.Gen.Genet. 235:213-220.

249. Rosenfeld, M.G. (1991) POU-domain transcription factors:pou-er-ful developmental regulators. Genes Dev. 5:897-907.

250. Rossant, J. and McBurney, M.W. (1982) The developmental potential of a euploid male teratocarcinoma cell line after blastocyst injection. J.Embryol.Exp.Morphol. 70:99-112.

251. Roy, A.L., Meisterernst, M., Pognonec, P. and Roeder, R.G. (1991) Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. nature 354:245-248.

252. Rudnicki, M.A. and McBurney, M.W. Cell culture methods and induction of differentiation of embryonal carcinoma cell lines. In: *Teratocarcinomas and embryonic stem cells:a practical approach*, edited by Robertson, E.J. Oxford: IRL Press, 1987, p. 19-49.

253. Ruppert, S., Wang, E.H. and Tjian, R. (1993) Cloning and expression of human TAF250:a TBP-associated factor implicated in cell-cycle regulation. nature 362:175-179.

254. Rushlow, C.A., Hogan, A., Pinchin, S.M., Howe, K.M., Lardelli, M. and Ish-Horowicz, D. (1989) The Drosophila hairy protein acts in both segmentation and bristle patterning and shows homology to N-myc. EMBO J. 8:3095-3103.

255. Ryder, K., Silver, S., DeLucia, A.L., Fanning, E. and Tegtmeyer, P. (1986) An altered DNA conformation in origin region 1 is a determinant for the binding of SV40 large T antigen. Cell 44:719-725.

256. Sakai, T., Ohtani, N., McGee, T.L., Robbins, P.D. and Dryja, T.P. (1991) Oncogenic germ-line mutations in Sp1 and ATF sites in the human retinoblastoma gene. nature 353:83-86.

257. Santoro, C., Mermod, N., Andrews, P.C. and Tjian, R. (1988) A family of human CCAAT-box-binding proteins active in transcription and DNA replication:cloning and expression of multiple cDNAs. nature 334:218-224.

258. Sauer, F. and Jäckle, H. (1993) Dimerization and the control of transcription by Krüppel. nature 364:454-457.

259. Sawadogo, M. and Sentenac, A. (1990) RNA polymerase B (II) and general transcription factors. Annu.Rev.Biochem. 59:711-754.

260. Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J.H.J., Chambon, P. and Egly, J.-M. (1993) DNA repair helicase:a component of BTF2 (TFIIH) basic transcription factor. Science 260:58-63.

261. Schild, C., Claret, F.-X., Wahli, W. and Wolffe, A.P. (1993) A nucleosome-dependent static loop potentiates estrogen-regulated transcription from the Xenopus vitellogenin B1 promoter in vitro. EMBO J. 12:423-433.

262. Schlissel, M.S. and Brown, D.D. (1984) The transcriptional regulation of Xenopus 5S RNA genes in chromatin:the roles of active stable transcription complexes and histone H1. Cell 37:903-913.

263. Schreck, R., Zorbas, H., Winnacker, E.-L. and Baeuerle, P.A. (1990) The NF-kappaB transcription factor induces DNA bending which is modulated by its 65-kd subunit. Nucl.AcidsRes. 18:6497-6502.

264. Sehgal, A., Patil, N. and Chao, M. (1988) A constitutive promoter directs expression of the nerve growth factor receptor gene. Mol.Cell.Biol. 8:3160-3167.

265. Sellers, J.W. and Struhl, K. (1989) Changing fos oncoprotein to a jun-dependent DNA-binding protein with GCN4 dimerization specificity by swapping 'leucine zippers'. nature 341:74-76.

266. Seth, A. and Papas, T.S. (1990) The c-ets-1 proto-oncogene has oncogenic activity and is positively autoregulated. Oncogene 5:1761-1768.

267. Seth, A., Watson, D.K., Blair, D.G. and Papas, T.S. (1989) c-ets-2 protooncogene has mitogenic and oncogenic activity. Proc.Natl.Acad.Sci.USA 86:7833-7837.

268. Seto, E., Lewis, B. and Shenk, T. (1993) Interaction between transcription factors Sp1 and YY1. nature 365:462-464.

269. Seto, E., Shi, Y. and Shenk, T. (1991) YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. nature 354:241-245.

270. Seto, E., Usheva, A., Zambetti, G.P., Momand, J., Horikoshi, N., Weinmann, R., Levine, A.J. and Shenk, T. (1992) Wild-type p53 binds to the TATA-binding protein and represses transcription. Proc.Natl.Acad.Sci.USA 89:12028-12032.

271. Shi, Y., Seto, E., Chang, L.-S. and Shenk, T. (1991) Transcriptional repression by YY1, a human GL1-Krüppel-related protein, and relief of repression by Adenovirus E1A protein. Cell 67:377-388.

272. Shoeman, R. and Traub, P. (1990) The in vitro DNA-binding properties of purified nuclear lamin proteins and vimentin. J.Biol.Chem. 265:9055-9061.

273. Shore, D. and Nasmyth, K. (1987) Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. Cell 51:721-732.

274. Shuey, D.J. and Parker, C.S. (1986) Bending of promoter DNA on binding of heat shock transcription factor. nature 323:459-461.

275. Shyam, E., Reddy, P. and Rao, V.N. (1990) Localization and modulation of the DNA-binding activity of the human c-ets-1 protooncogene. Cancer Res. 50:5013-5016.

276. Simon, M.C., Fisch, T.M., Benecke, B.J., Nevins, J.R. and Heintz, N. (1988) Definition of multiple, functionally distinct TATA elements, one of which is a target in the hsp70 promoter for E1A regulation. Cell 52:723-729.

277. Simpson, P. (1990) Lateral inhibition and the developemnt of the sensory bristles of the adult PNS in Drosophila. Development 109:509-515.

278. Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.-M., Lovell-Badge, R. and Goodfellow, P.N. (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. nature 346:240-244.

279. Singh, H., LeBowitz, J.H., Baldwin, A.S. and Sharp, P.A. (1988) Molecular cloning of an enhancer binding protein:isolation by screening of an expression library with a recognition site DNA. Cell 52:415-423.

280. Sleigh, M.J. (1986) A nonchromatographic assay for expression of the chloramphenicol acetyltransferase gene in eukaryotic cells. Anal.Biochem. 156:251-256.

281. Smale, S.T. and Baltimore, D. (1989) The "initiator" as a transcription control element. Cell 57:103-113.

282. Sollner-Webb, B. (1988) Surprises in polymerase III transcription. Cell 52:153-154.

283. Speck, N.A. and Baltimore, D. (1987) Six distinct nuclear factors interact with the 75-base-pair repeat of the moloney murine leukemia virus enhancer. Mol.Cell.Biol. 7:1101-1110.

284. St-Arnaud, R. and Moir, J.M. (1993) Wnt-1-inducing factor-1:a novel G/C box-binding transcription factor regulating the expression of Wnt-1 during neuroectodermal differentiation. Mol.Cell.Biol. 13:1590-1598.

285. Staben, C. and Yanofsky, C. (1991) Neurospora crassa a mating-type region. Proc.Natl.Acad.Sci.USA 87:4917-4921.

286. Stanway, C., Mellor, J., Ogden, J.E., Kingsman, A.J. and Kingsman, S.M. (1987) The UAS of the yeast Pgk gene contains functionally distinct domains. Nucl.AcidsRes. 15:6855-6873.

287. Stein, B., Cogswell, P.C. and Baldwin, A.S.,Jr. (1993) Functional and physical associations between NF-kappaB and C/EBP family members:a rel domain-bZIP interaction. Mol.Cell.Biol. 13:3964-3974.

288. Stein, G.S., Lian, J.B., Dworetzky, S.I., Owen, T.A., Bortell, R., Bidwell, J.P. and van Wijnen, A.J. (1991) Regulation of transcription-factor activity during growth and differentiation: involvement of the nuclear matrix in concentration and localization of promoter binding proteins. J.Cell.Biochem. 47:300-305.

289. Stief, A., Winter, D.M., Strätling, W.H. and Sippel, A.E. (1989) A nuclear DNA attachment element mediates elevated and position-independent gene activity. nature 341:343-345.

290. Struhl, K. (1989) Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. Trends Biochem.Sci. 14:137-140.

291. Stuurman, N., Meijne, A.M.I., van der Pol, A.J., deJong, L., van Driel, R. and van Renswoude, J. (1990) The nuclear matrix from cells of different origin. J.Biol.Chem. 265:5460-5465.

292. Sun, X.-H., Copeland, N.G., Jenkins, N.A. and Baltimore, D. (1991) Id proteins Id1 and Id2 selectively inhibit DNA binding by one class of helix-loop-helix proteins. Mol.Cell.Biol. 11:5603-5611.

293. Sundseth, R. and Hansen, U. (1992) Activation of RNA polymerase II transcription by the specific DNA-binding protein LSF. J.Biol.Chem. 267:7845-7855.

294. Takagi, N. and Sasaki, M. (1975) Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. nature 256:640-642.

295. Tamaru, M., Nagao, Y., Taira, M., Tatibana, M., Masamune, Y. and Nakanishi, Y. (1990) Selective activation of testis-specific genes in cultured rat spermatogenic cells. Biochim.Biophys.Acta 1049:331-338.

296. Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.F., Weintraub, H. and Lassar, A.B. (1988) MyoD:a nuclear phosphoprotein requiring a myc homology region to convert fibroblasts to myoblasts. Science 242:405-411.

297. Tartof, K.D. and Hobbs, C.A. (1987) Improved media for growing plasmid and cosmid clones. Focus 9.2:12.

298. Taylor, I.C.A., Workman, J.L., Schuetz, T.J. and Kingston, R.E. (1991) Facilitated binding of GAL4 and heat shock factor to nucleosomal templates:differential function of DNA-binding domains. Genes Dev. 5:1285-1298.

299. Tazi, J. and Bird, A. (1990) Alternative chromatin structure at CpG islands. Cell 60:909-920.

300. Thayer, M.J. and Weintraub, H. (1993) A cellular factor stimulates the DNA-binding activity of myoD and E47. Proc.Natl.Acad.Sci.USA 90:6483-6487.

301. Thompson, C.B., Wang, C., Ho, I.-C., Bohjanen, P.R., Petryniak, B., June, C.H., Meisfeldt, S., Zhang, L., Nabel, G.J., Karpinski, B. and Leiden, J.M. (1992) cis-acting sequences required for inducible interleukin-2 enhancer function bind a novel ets-related protein, Elf-1. Mol.Cell.Biol. 12:1043-1053.

302. Thompson, C.C., Brown, T.A. and McKnight, S.L. (1991) Convergence of Ets- and notch- related structural motifs in a heteromeric DNA binding complex. Science 253:762-768.

303. Timmers, H.Th.M. and Sharp, P.A. (1991) The mammalian TFIID protein is present in two functionally distinct complexes. Genes Dev. 5:1946-1956.

304. Tjian, R. (1981) T antigen binding and the control of SV40 gene expression. Cell 26:1-2.

305. Travers, A.A. (1992) The reprogramming of transcriptional competence. Cell 69:573-575.

306. Travis, A., Amsterdam, A., Berlanger, C. and Grosschedl, R. (1991) LEF-1, a gene encoding a lymphoid-specific protein, with an HMG domain, regulates T-cell receptor α enhancer function. Genes Dev. 5:880-894.

307. Treacy, M.N., He, X. and Rosenfeld, M.G. (1991) I-POU:a POU-domain protein that inhibits neuron-specific gene activation. nature 350:577-584.

308. Treisman, R. and Maniatis, T. (1985) SV40 enhancer increases the number of RNA pol.II molecules on linked DNA. nature 315:72-75.

309. Tyree, C.M., George, C.P., Lira-DeVito, L.M., Wampler, S.L., dahmus, M.E., Zawel, L. and Kadonaga, J.T. (1993) Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. Genes Dev. 7:1254-1265.

310. Ullman, K.S., Flanagan, W.M., Edwards, C.A. and Crabtree, G.R. (1991) Activation of early gene expression in T lymphocytes by Oct-1 and an inducible protein, OAP40. Science 254:558-562.

311. van de Wetering, M., Oosterwegel, M., Dooijes, D. and Clevers, H. (1991) Identification and cloning of TCF-1, a lymphocyte-specific transcription factor containing a sequence-specific HMG-box. EMBO J. 10:123-132.

312. van Driel, R., Humbel, B. and de Jong, L. (1991) The nucleus:a black box being opened. J.Cell.Biochem. 47:311-316.

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

314. Verdin, E., Paras, P.,Jr. and Van Lint, C. (1993) Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. EMBO J. 12:3249-3259.

315. Verrijzer, C.P., van Oosterhout, J.A.W.M., van Weperen, W.W. and van der Vliet, P.C. (1991) POU proteins bend DNA via the POU-specific domain. EMBO J. 10:3007-3014.

316. Vidali, G., Boffa, L.C., Bradbury, E.M. and Allfrey, V.G. (1978) Butyrate suppression of histone deacetylation leads to accumulation of multi-acetylated forms of histones H3 and H4 and increased DNase 1 sensitivity of the associated DNA sequences. Proc.Natl.Acad.Sci.USA 75:2239-2243.

317. Vignais, M.-L. and Sentenac, A. (1989) Asymmetric DNA bending induced by the yeast multifunctional factor TUF. J.Biol.Chem. 264:8463-8466.

318. Villares, R. and Cabrera, C.V. (1987) The achaete-scute gene complex of D.melanogaster:conserved domains in a subset of genes required for neurogenesis and their homology to myc. Cell 50:415-424.

319. Vinson, C.R., LaMarco, K.L., Johnson, P.F., Landschulz, W.H. and McKnight, S.L. (1988) In situ detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. Genes Dev. 2:801-806.

320. von Kries, J.P., Buhrmester, H. and Strätling, W.H. (1991) A matrix/scaffold attachment region binding protein: identification, purification, and mode of binding. Cell 64:123-135.

321. Voss, J.W., Wilson, L. and Rosenfeld, M.G. (1991) POU-domain proteins Pit-1 and Oct-1 interact to form a heteromeric complex and can cooperate to induce expression of the prolactin promoter. Genes Dev. 5:1309-1320.

322. Walker, J., Chen, T.A., Sterner, R., Berger, M., Winston, F. and Allfrey, V.G. (1990) Affinity chromatography of mammalian and yeast nucleosomes. J.Biol.Chem. 265:5736-5746.

323. Walker, J.M., Gooderham, K., Hastings, J.R.B., Mayes, E. and Johns, E.W. (1980) The primary structures of non-histone chromosomal proteins HMG 1 and 2. FEBS Lett. 122:264-270.

324. Wasylyk, B., Wasylyk, C., Flores, P., Begue, A., Leprince, D. and Stehelin, D. (1990) The c-ets proto-oncogenes encode transcription factors that cooperate with c-Fos and c-Jun for transcriptional activation. nature 346:191-193.

325. Wasylyk, C., Kerckaert, J.-P. and Wasylyk, B. (1992) A novel modulator domain of Ets transcription factors. Genes Dev. 6:965-974.

326. Weber, F. and Schaffer, W. (1985) SV40 enhancer increases RNA polymerase density within the linked gene. nature 315:75-77.

327. Wefald, F.C., Devlin, B.H. and Williams, R.S. (1990) Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. nature 344:260-262.

328. Weinzierl, R.O.J., Dynlacht, B.D. and Tjian, R. (1993) Largest subunit of Drosophila transcription factor IID directs assembly of a complex containing TBP and a coactivator. nature 362:511-517.

329. Weis, L. and Reinberg, D. (1992) Transcription by RNA polymerase II:initiator-directed formation of transcription-competent complexes. FASEB J. 6:3300-3309.

330. West, J.D., Frels, W.I., Chapman, V.M. and Papaioannou, V.E. (1977) Preferential expression of the maternally derived X chromosome in the mouse yolk sac. Cell 12:873-882.

331. Wiley, S.R., Krause, R.J. and Mertz, J.E. (1992) Functional binding of the TATA box binding-component of transcription factor TFIID to the -30 region of TATA-less promoters. Proc.Natl.Acad.Sci.USA 89:5814-5818.

332. Williams, T. and Tjian, R. (1991) Analysis of the DNA-binding and activation properties of the human transcription factor AP-2. Genes Dev. 5:670-682.

333. Wright, W.E., Sassoon, D.A. and Lin, V.K. (1989) Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. Cell 56:607-617.

334. Xin, J., Cowie, A., Lachance, P. and Hassell, J.A. (1992) Molecular cloning and characterization of PEA3, a new member of the Ets oncogene family that is differentially expressed in mouse embryonic cells. Genes Dev. 6:481-496.

335. Yang, T.P., Singer-Sam, J., Flores, J.C. and Riggs, A. (1988) DNA binding factors for the CpG-rich island containing the promoter of the human X-linked PGK gene. Som.Cell and Mol.Gen. 14:461-472.

336. Yuan, J., Simos, G., Blobel, G. and Georgatos, S.D. (1991) Binding of lamin A to polynucleosomes. J.Biol.Chem. 266:9211-9215.

337. Zawel, L. and Reinberg, D. (1992) Advances in RNA polymerase II transcription. Curr.Opin.Cell Biol. 4:488-495.

338. Zenzie-Gregory, B., Khachi, A., Garraway, I.P. and Smale, S.T. (1993) Mechanism of initiator-mediated transcription:evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. Mol.Cell.Biol. 13:3841-3849.

339. Zhou, Q., Lieberman, P.M., Boyer, T.G. and Berk, A.J. (1992) Holo-TFIID supports transcriptional stimulation by diverse activators and from a TATA-less promoter. Genes Dev. 6:1964-1974.

340. Zhu, H., Roy, A.L., Roeder, R.G. and Prywes, R. (1991) Serum response factor affects preinitiation complex-formation by TFIID in vitro. New Biologist 3:455-464.

341. Ziff, E.B. (1990) Transcription factors:a new family gathers at the cAMP response site. Trends Genet. 6:69-72.