THE CONTROL OF RETROVIRAL GENE EXPRESSION
IN PLURIPOTENTIAL EMBRYONIC STEM CELLS

A thesis presented by

Victoria Emma Prince

for
the degree of
Doctor of Philosophy

at
University College
University of London
in collaboration with the
National Institute for Medical Research
Mill Hill
London
(January 1991)
For my family, without whom none of this would have been possible.
ABSTRACT

The Moloney murine leukaemia virus (Mo-MuLV) and its close relative the Moloney murine sarcoma virus (Mo-MSV) are retroviruses incapable of productively infecting early embryonic cells. The viral long terminal repeat (LTR) sequences are unable to direct transcription in embryonic cells but are highly active in their differentiated derivatives; hence these viruses provide a useful probe for investigating transcriptional control in stem cells.

Protein extracts of F9 embryonal carcinoma stem cells and their differentiated derivatives were used in gel retardation experiments with Mo-MSV LTR sequences. The interactions were compared to those with sequences from the LTRs of mutant derivatives of Mo-MuLV which are transcribed in early embryonic cells. Two mutant viruses, the myeloproliferative sarcoma virus (MPSV) and its derivative PCMV, share a point mutation at -166 base-pairs (bp) with respect to the transcriptional start, which generates a consensus binding site for the well characterized transcription factor Spl. The PCMV sequence around -166 interacted with a protein of equivalent immuno-reactivity, DNA-binding specificity and electrophoretic mobility to purified human Spl protein. The wild-type sequence had 5-fold lower affinity for this protein. Consistent with this observation, transient transfection experiments showed that the point mutation at -166 increased transcription 6-fold.

Gel retardation experiments showed that the wild-type sequence around -345 bp interacted with an EC cell-specific protein. The protein had 10-fold lower affinity for the equivalent PCMV sequence, which has a point mutation at -345. As PCMV can be transcribed in EC cells the protein is an excellent candidate for a repressor of transcription. However, the mutation did not increase transcription in transient transfection experiments.

The gene products of the adenovirus early gene, E1A, down-regulate papovavirus transcription. Co-transfection experiments showed that E1A also down-regulates transcription of Mo-MSV in an enhancer dependent fashion.
CHAPTER 1 INTRODUCTION

1.1 GENERAL INTRODUCTION

1.2 EARLY MOUSE DEVELOPMENT

1.2.1. In vitro model systems of early mouse development.

1.2.2. Gene expression during preimplantation development.

1.3 TRANSCRIPTION

1.3.1. Control of initiation of transcription.

1.3.2. Transcription factors.

1.3.3. Transcription factor Spl.

1.3.4. The CCAAT box-binding family.

1.3.5. Negative regulation of transcription.

1.4 VIRAL TRANSCRIPTION

1.4.1. Simian virus 40.

1.4.2. Polyoma virus.

1.4.3. Adenovirus.

1.4.4. Viral transcription in embryonic stem cells.

1.5 RETROVIRUSES

1.5.1. Life cycle.

1.5.2. Moloney murine leukaemia virus.

1.5.3. Transcriptional control of Mo-MuLV.

1.5.4. The non-permissivity of stem cells for Mo-MuLV.

1.6 AIMS OF THE PROJECT

CHAPTER 2 MATERIALS AND METHODS

2.1 GENERAL METHODS

2.1.1. Safety.

2.1.2. Sterilization.
2.1.3. Extractions and precipitations.
2.1.4. Preparation of dialysis tubing.

2.2 SYNTHETIC OLIGONUCLEOTIDES

2.3 RESTRICTION ENDOUCLEASE DIGESTION

2.3.1. Diagnostic digestions.
2.3.2. Preparative digestions.

2.4 SUBCLONING

2.4.1. Deposphorylation of DNA.
2.4.2. Ligation.
2.4.3. Preparation of competent cells.
2.4.4. Transformation.
2.4.5. Identification of positive clones.

2.5 SITE-DIRECTED MUTAGENESIS

2.5.1. Growth of Uracil-containing phagemids.
2.5.2. Extraction of phagemid DNA.
2.5.3. Synthesis of the mutagenic strand.
2.5.4. Transformation into MV1190 and analysis of transformants.

2.6 PREPARATION OF PLASMID DNA

2.7 SEQUENCING

2.8 END LABELLING

2.8.1. Labelling using T4 polynucleotide kinase.
2.8.2. Labelling using DNA polymerase I (Klenow fragment).
2.8.3. Measurement of incorporation of \( ^{32}P \).

2.9 TISSUE CULTURE

2.10 F9 CELL DIFFERENTIATION \textit{IN VITRO}

2.11 ANTI-SSEA-1 STAINING

2.12 TRANSFECTION

2.13 CAT ASSAYS

2.14 ASSAY FOR \( \beta \)-GALACTOSIDASE PRODUCTION

2.15 RNA EXTRACTION

2.16 WHOLE-CELL EXTRACTS

2.17 \textit{IN VITRO} TRANSCRIPTION

2.18 PRIMER EXTENSION

2.19 RNASE PROTECTION
CHAPTER 3 A COMPARISON OF RETROVIRUS LTR-BINDING PROTEINS IN EC CELLS AND THEIR DIFFERENTIATED DERIVATIVES

3.1 INTRODUCTION

3.2 CONTROLS FOR INTEGRITY AND COMPARABILITY OF WHOLE-CELL EXTRACTS

3.3 GEL RETARDATION EXPERIMENTS USING RESTRICTION FRAGMENTS OF THE Mo-MSV LTR AS PROBES

3.4 GEL RETARDATION EXPERIMENTS USING SYNTHETIC OLGONUCLEOTIDES

3.5 DISCUSSION

CHAPTER 4 ANALYSIS OF THE -345 MUTATION OCCURRING IN THE PCMV MUTANT RETROVIRUS

4.1 INTRODUCTION

4.2 COMPLEX b INVOLVES THE SEQUENCE BETWEEN -372 AND -329

4.3 THE PCMV SEQUENCE AROUND -345 FORMS COMPLEX b WITH APPROXIMATELY 10-FOLD LOWER AFFINITY THAN THE EQUIVALENT WILD-TYPE SEQUENCE

4.4 OFF-RATE ANALYSIS CONFIRMS THAT COMPLEX b FORMS WITH HIGHER AFFINITY ON THE WILD-TYPE SEQUENCE AROUND -345 THAN ON THE CORRESPONDING PCMV SEQUENCE

4.5 A SEARCH FOR SEQUENCES HOMOLOGOUS TO THE EC-R BINDING SITE

4.6 EC-R DOES NOT RECOGNIZE PEA-2 OR PEBP5 BINDING SITES

4.7 HEPARIN-SEPHAROSE COLUMN CHROMATOGRAPHY AFFORDS 20-FOLD PURIFICATION OF EC-R
4.8 Wheat germ agglutinin column chromatography shows that EC-R is not glycosylated.

4.9 FPLC analysis shows that EC-R has a native molecular weight of about 60kd.

4.10 A single cycle of affinity chromatography affords a further 10-fold purification of EC-R.

4.11 Discussion

Chapter 5 Analysis of the point mutation at -166 occurring in mutant retroviruses capable of being transcribed in embryonal carcinoma stem cells

5.1 Introduction

5.2 The PCMV sequence around -166 competes for complexes A-C more efficiently than the corresponding Mo-MSV sequence.

5.3 The PCMV sequence from -176 to -157 has approximately 5-fold higher affinity for complexes A, B and C than the equivalent Mo-MSV sequence.

5.4 Complexes A-C are competed for by an HSV tk gene promoter fragment that includes two Sp1 binding sites.

5.5 Complex A is recognized by anti-Sp1 antibodies.

5.6 Purified Sp1 protein from HELA cells forms a complex with the PCMV oligonucleotide of equivalent mobility and behaviour in competition experiments to complex A.

5.7 Western blot analysis shows that F9 EC cells contain a protein of equivalent size and immuno-reactivity to HELA cell Sp1.

5.8 Wheat germ agglutinin column chromatography suggests that the protein components of complexes A-C are differentially glycosylated.

5.9 Heparin-Sepharose ion-exchange column chromatography shows that the protein component of complex B can be separated from those of complexes A and C.
CHAPTER 6 IN VIVO ANALYSES OF THE ROLE OF POINT MUTATIONS IN THE TRANSCRIPTION OF MUTANT RETROVIRUSES

6.1 TRANSIENT TRANSFECTION ANALYSES OF THE EFFECTS OF POINT MUTATIONS ON THE TRANSCRIPTION OF Mo-MSV IN EC CELLS

6.1.1. Introduction of point mutations to pMSVcat.

6.1.2. CAT assays.

6.1.3. The Mo-MSV enhancer is active in fibroblasts but inactive in EC cells.

6.1.4. The -166 mutation increases Mo-MSV transcription 6-fold in transient transfection experiments.

6.1.5. The -345 mutation does not increase Mo-MSV transcription in transient transfection experiments.

6.1.6. Discussion.

6.2 REGULATION OF RETROVIRUS TRANSCRIPTION BY THE ADENO VIRUS E1A GENE PRODUCT

6.2.1. The adenovirus E1A gene product down-regulates transcription from the Mo-MSV LTR.

6.2.2. Discussion.

CHAPTER 7 FINAL DISCUSSION

7.1 SUMMARY OF CONCLUSIONS

7.2 FURTHER CONSIDERATION OF THE ANALYSIS OF LTR BINDING PROTEINS

7.2.1. The -345 bp binding protein.

7.2.2. The -166 bp binding proteins.

7.3 FURTHER CONSIDERATION OF THE RESULTS OF TRANSIENT TRANSFECTION EXPERIMENTS

7.4 A MODEL TO EXPLAIN TRANSCRIPTIONAL ACTIVITY OF MUTANT RETROVIRUSES IN EC CELLS

7.5 PROSPECTS FOR FUTURE STUDY

ACKNOWLEDGEMENTS

REFERENCES
LIST OF FIGURES

Page no.  
19  Fig. 1.1. Early murine development.  
54  Fig. 1.2. Structure of the Moloney murine leukaemia virus genome.  
56  Fig. 1.3. Retroviral life cycle.  
108 Fig. 3.1. Anti-SSEA1 staining of F9 EC cells and their differentiated derivatives.  
110 Fig. 3.2. Whole-cell extracts can support transcription in vitro.  
113 Fig. 3.3. F9 EC and parietal endoderm-like extracts contain equivalent concentrations of a non differentiation-regulated factor.  
117 Fig. 3.4. Sequence of the Mo-MSV 5' LTR.  
120 Fig. 3.5. Non-differentiation regulated complexes form with the Mo-MSV LTR between -252 and -152 bp with respect to the transcriptional start.  
122 Fig. 3.6. An EC cell-specific complex forms on the Mo-MSV LTR between -372 and -280 bp with respect to the transcriptional start.  
125 Fig. 3.7. An EC cell specific complex forms with the Mo-MSV sequence from -363 to -333 with higher affinity than with the corresponding PCMV sequence.  
130 Fig. 3.8. Gel retardation analysis using the NF-1 WT oligonucleotide.  
132 Fig. 3.9. The -176 to -157 sequence forms three specific complexes with a variety of extracts.  
138 Fig. 4.1. Competition analysis of complexes a and b formed on the Mo-MSV sequence from -372 to -280 using DNA fragments.  
141 Fig. 4.2. Competition analysis of complexes a and b formed on the Mo-MSV sequence from -372 to -280 using synthetic oligonucleotides.
Fig. 4.3. Off-rate analysis of complex b when formed on the Mo-MSV or PCMV sequences around -345.

Fig. 4.4. Competition of the EC-R binding site with the polyomavirus A element.

Fig. 4.5. Partial purification of EC-R by heparin-Sepharose column chromatography.

Fig. 4.6. Wheat germ agglutinin column chromatography shows that EC-R is not glycosylated.

Fig. 4.7. Estimation of the native molecular weight of EC-R using FPLC.

Fig. 4.8. Further purification of EC-R by affinity chromatography.

Fig. 5.1. Competition analysis of the complexes formed on the PCMV (Spl WT) probe.

Fig. 5.2. Competition analysis of the complexes formed on the PCMV probe by a titration of unlabelled Mo-MSV and PCMV probes.

Fig. 5.3. Complexes A-C are competed for by an HSV tk gene promoter fragment containing two Spl binding sites.

Fig. 5.4. The protein involved in complex A is recognized by anti-Spl antibody.

Fig. 5.5. Complex A comigrates with a complex formed between purified Spl protein and the PCMV oligonucleotide.

Fig. 5.6. A murine equivalent of human Spl protein is detectable by Western blotting.

Fig. 5.7. Wheat germ agglutinin column chromatography shows that the protein components of complexes B and C are under-glycosylated with respect to the Spl-like protein of complex A.

Fig. 5.8. Heparin-Sepharose column chromatography separates complex B from complexes A and C.
195 Fig. 6.1. Strategy for cloning point mutations into pMSVcat.

198 Fig. 6.2. Formation of covalently closed circular DNA during formation of the mutagenic strand.

201 Fig. 6.3. Sequencing of wild-type and mutated sequences around -166.

203 Fig. 6.4. Sequencing of wild-type and mutated sequences around -345.

206 Fig. 6.5. CAT assay using purified enzyme.

210 Fig. 6.6. The Mo-MSV enhancer is active in differentiated cells.

213 Fig. 6.7. The -166 mutation increases transcription in F9 EC cells by 6-fold.

216 Fig. 6.8. The -166 mutation increases transcription in SV3T3 cells by 2-fold.

219 Fig. 6.9. Primer extension analysis of RNA transcripts from pMSVcat and pMSV(-166)cat in SV3T3 cells.

221 Fig. 6.10. RNAse protection analysis of RNA transcripts from pMSVcat and pMSV(-166)cat in F9 EC cells.

224 Fig. 6.11. pMSV(-166)cat and pMSV(-166,-345)cat have equivalent transcriptional activity in F9 EC cells.

226 Fig. 6.12. Wild-type and mutant Mo-MSV LTR sequences direct equivalent levels of transcription in PCC4 EC cells.

233 Fig. 6.13. Western blot analysis shows that E1A protein is expressed from the pRSV_E1A expression plasmid on transfection of SV3T3 cells.

238 Fig. 6.14. E1A down-regulates transcription from the Mo-MSV LTR in an enhancer dependent, and point mutation independent fashion.
LIST OF TABLES

52  TABLE 1.1 The functions of each region of the Mo-MuLV genome

77  TABLE 2.1 Electrophoresis buffers

78  TABLE 2.2 Conditions for restriction endonuclease digestions

80  TABLE 2.3 E. coli strains

83  TABLE 2.4 Media used for growth of E. coli

84  TABLE 2.5 Plasmids used - sources and descriptions

90  TABLE 2.6 Sequencing buffers

200 TABLE 6.1 Mutagenesis of pMSVcat at -166 and reversion of the "control phagemid"

215 TABLE 6.2 Relative CAT expression from pMSV(-166)cat and pMSVcat in F9 EC cells

235 TABLE 6.3 Expression from pMSVcat is down-regulated by the E1A gene products

236 TABLE 6.4 E1A down-regulation of Mo-MSV transcription is enhancer dependent and point-mutation independent
ABBREVIATIONS

A  absorbance; adenine
Ad  adenovirus
ATF  activating transcription factor
ATP  adenosine 5’-triphosphate
bp  base-pairs
BHI  brain heart infusion
Bis  N,N’-methylenebisacrylamide
BSA  bovine serum albumin, fraction V
C  cytosine
CaCl₂  calcium chloride
cAMP  cyclic adenosine-3’,5’ monophosphate
cat  chloramphenicol acetyltransferase (gene)
CAT  chloramphenicol acetyltransferase (protein)
CIP  calf intestinal alkaline phosphatase
cpm  counts per minute
CRE  cAMP response element
CsCl  caesium chloride
CTP  cytidine 5’-triphosphate
dATP  2’-deoxyadenosine 5’-triphosphate
dCTP  2’-deoxycytidine 5’-triphosphate
dGTP  2’-deoxyguanosine 5’-triphosphate
dNTP  2’-deoxynucleoside 5’-triphosphate
dTTP  2’-deoxythymidine 5’-triphosphate
ddATP  2’,3’-dideoxyadenosine 5’-triphosphate
ddCTP  2’,3’-dideoxycytidine 5’-triphosphate
ddGTP  2’,3’-dideoxyguanosine 5’-triphosphate
ddNTP  2’,3’-dideoxynucleoside 5’-triphosphate
ddTTP  2’,3’-dideoxythymidine 5’-triphosphate
dGTP  2’-deoxyguanosine 5’-triphosphate
DEPC  diethyl pyrocarbonate
dH₂O  distilled water
DMEM  Dulbecco’s modification of Eagle’s medium
DMS  dimethylsulphate
DMSO  dimethyl sulphoxide
DNA  deoxyribonucleic acid
DNAse  deoxyribonuclease
dpc  days post coitum
ds  double-stranded
DTT  dithiothreitol
EC  embryonal carcinoma
E. coli  Escherichia coli
EDTA  diaminooethane tetraacetic acid
ES  embryonic stem
EtBr  ethidium bromide
EtOH  ethanol
FCS  foetal calf serum
FPLC  fast protein liquid chromatography
Fr-MuLV  Friend murine leukaemia virus
G  guanine
GR  glucocorticoid receptor
GRE  glucocorticoid response element
GTP  guanosine 5'-triphosphate
h  hours
HCl  hydrochloric acid
Hepes  N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
HLH  helix-loop-helix
hnRNA  heterogeneous nuclear RNA
HPRT  hypoxanthine phosphoribosyl transferase
hsp  heat shock protein
HSV  Herpes Simplex virus
ICM  inner cell mass
IPTG  isopropyl-β-D-thiogalactoside
kb  kilobase
KCl  potassium chloride
kd  kilodalton
KH₂PO₄  potassium dihydrogen orthophosphate
LGT  low gelling temperature
LTR  long terminal repeat
μCi  microcuries
μg  microgramme
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>μl</td>
<td>microliters</td>
</tr>
<tr>
<td>μM</td>
<td>micromolar</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>mg</td>
<td>milligramme</td>
</tr>
<tr>
<td>MHC</td>
<td>major histocompatibility complex</td>
</tr>
<tr>
<td>min</td>
<td>minutes</td>
</tr>
<tr>
<td>ml</td>
<td>millilitre</td>
</tr>
<tr>
<td>MLP</td>
<td>major late promoter</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-(N-morpholino) propanesulphonic acid</td>
</tr>
<tr>
<td>MPSV</td>
<td>myeloproliferative sarcoma virus</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>Mo-MSV</td>
<td>Moloney murine sarcoma virus</td>
</tr>
<tr>
<td>Mo-MuLV</td>
<td>Moloney murine leukaemia virus</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium chloride</td>
</tr>
<tr>
<td>Na₂HPO₄</td>
<td>disodium hydrogen orthophosphate</td>
</tr>
<tr>
<td>NaOH</td>
<td>sodium hydroxide</td>
</tr>
<tr>
<td>NF-1</td>
<td>nuclear factor-1</td>
</tr>
<tr>
<td>NP40</td>
<td>nonidet P40</td>
</tr>
<tr>
<td>NRE</td>
<td>negative regulatory element</td>
</tr>
<tr>
<td>NTP</td>
<td>nucleoside 5’-triphosphate</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline (Dulbecco’s solution A)</td>
</tr>
<tr>
<td>PCMV</td>
<td>PCC4 cell-passaged neomycin resistant MPSV</td>
</tr>
<tr>
<td>PE</td>
<td>parietal endoderm</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PIPES</td>
<td>piperazine N,N'-bis(2-ethane-sulphonic acid)</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulfonyl fluoride</td>
</tr>
<tr>
<td>Py</td>
<td>polyomavirus</td>
</tr>
<tr>
<td>RA</td>
<td>retinoic acid</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>RSV</td>
<td>Rous sarcoma virus</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>ss</td>
<td>single-stranded</td>
</tr>
<tr>
<td>SSEA</td>
<td>stage specific embryonic antigen</td>
</tr>
<tr>
<td>SV40</td>
<td>Simian virus 40</td>
</tr>
<tr>
<td>T</td>
<td>thymine</td>
</tr>
<tr>
<td>TE</td>
<td>trophectoderm; 10mM Tris(pH 8.0), 1mM EDTA</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-tetramethylethylenediamine</td>
</tr>
<tr>
<td>TF</td>
<td>transcription factor</td>
</tr>
<tr>
<td>TPA</td>
<td>12-tetradecanoylphorbol-13-acetate</td>
</tr>
<tr>
<td>t-PA</td>
<td>tissue plasminogen activator</td>
</tr>
<tr>
<td>Tris</td>
<td>tris (hydroxymethyl) methylamine</td>
</tr>
<tr>
<td>Tween 20</td>
<td>polyoxyethylene sorbitan monolaurate</td>
</tr>
<tr>
<td>U</td>
<td>uracil</td>
</tr>
<tr>
<td>UTP</td>
<td>uridine 5’-triphosphate</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>VSV</td>
<td>vesicular stomatitis virus</td>
</tr>
<tr>
<td>WGA</td>
<td>wheat germ agglutinin</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION
1.1 GENERAL INTRODUCTION

One of the fundamental problems of modern biology is to understand the mechanisms controlling embryonic development. Development of a single fertilized egg cell into a multicellular animal requires control at the level of spatial organization and also at the level of cellular differentiation. When a cell differentiates it begins to synthesize a different array of proteins to those it or its ancestors previously made, thus allowing it to take on a new and more specialized function. Stem cells are unusual in having a relatively undifferentiated phenotype and being capable of giving rise both to more stem cells and to cells of a differentiated phenotype. They occur not only in the developing embryo but also in the adult organism to allow continued generation of new cells. As one of the most important points at which gene expression is controlled is at the level of messenger RNA (mRNA) transcription, the control of transcription in stem cells is likely to differ from that in their more differentiated derivatives. Thus one route towards the further understanding of embryonic development is to study the control of transcription in stem cells.

1.2 EARLY MOUSE DEVELOPMENT

The embryonic development of the mouse begins when the egg is fertilized by the sperm and is completed after nineteen or twenty days depending on the mouse strain. Under optimum conditions ovulation will occur every four days in response to a surge in the level of luteinizing hormone. This causes eight to ten oocytes surrounded by the zona and follicle cells to be swept into the infundibulum of the oviduct over a period of two to three hours. Sperm are released into the reproductive tract and age for about one hour before penetrating the follicle and zona. For successful fertilization the acrosome reaction must occur at this point releasing hydrolytic enzymes. Fusion of sperm head and egg membrane then triggers fertilization.

Prior to fertilization the oocyte begins the first division of
meiosis, but is arrested during the diplotene stage of prophase to allow the egg to grow and mature in the resting or dictyate stage which lasts about 15 days. During this stage maternal mRNA transcripts accumulate within the oocyte, but are not translated until after fertilization. The first meiotic division is completed and one diploid set of chromosomes is segregated outside the egg in the first polar body. The second meiotic division initiates but does not complete until triggered by sperm entry. As the maternal pronucleus forms, the other haploid set of chromosomes is eliminated in the second polar body, and the sperm chromatin decondenses to allow formation of the male pronucleus.

DNA replication occurs as the maternal and paternal chromosomes migrate to the centre of the egg and the first cleavage occurs at about 24 hours after fertilization. Up to the eight cell stage the cells remain equipotent, each capable of giving rise to a mouse (Kelly, 1977). By the sixteen cell stage compaction has occurred to produce the morula. The cells now have clear differences, each cell has an inside and an outside, and two distinct lineages are already beginning to develop. The inner cell mass (ICM) will give rise to embryonic tissue and the trophectoderm (TE) will form the placenta. This differentiation process begins with compaction and ends with formation of a fluid filled blastocyst consisting of twenty ICM cells and forty-four TE cells (see Fig. 1.1.)

At about 4 days post-coitum (dpc) the second round of differentiation events occurs. The ICM differentiates into two separate cell types; an epithelial layer known as the primitive endoderm which forms on the free surface of the ICM, and the epiblast or primitive ectoderm, which will give rise to the embryo. Injection chimaera experiments using lineage markers show that the primitive endoderm cells do not colonize foetal cells but only the extraembryonic parietal and visceral endoderm of the yolk sacs (Gardner, 1982; 1983). Similar experiments show the primitive ectoderm to be pluripotent; it gives rise to ectodermal, mesodermal and endodermal tissues of the embryo (Gardner and Rossant, 1979). The TE becomes regionally specialized to form both the mural trophectoderm which is in contact with the blastocoel and the polar
Fig. 1.1. Early murine development. The early stages of murine development from fertilization to 7 days post-coitum (dpc) are illustrated. (Taken from Jackson, 1989.)
zona pellucida

♀ pronucleus
polar bodies

♂ pronucleus
~8 hours p.c.
early cleavage: cells loosely associated
~24 hours p.c.
~40 hours p.c.

8 cell embryo
~48 hours p.c.

compacted morula
tight junction formed
~54 hours p.c.
early blastocyst
64 cells
~96 hours p.c.

~108 hours p.c.
hatching blastocyst

polar
trophectoderm
epiblast
mural trophectoderm
4.5 dpc

ectoplacental cone

parietal endoderm
Primitive Endoderm
epiblast (primitive ectoderm)
giant cells
5.5 dpc

extraembryonic ectoderm
embryonic ectoderm
proamniotic cavity
visceral endoderm (embryonic)

amniotic fold
primitive streak
A
P
mesoderm
parietal endoderm
visceral endoderm

6 dpc (egg cylinder)
7 dpc
trophectoderm above the ICM. The mural cells do not divide any further but continue DNA replication to form polyploid giant cells.

At approximately 4.5 dpc the blastocyst hatches from the zona-pelludica and implants into the uterine wall. The mural trophoblast contacts the uterine wall and penetrates the epithelial layer, inducing a proliferation of uterine cells at this site. By 5.5 dpc the egg cylinder has formed as a result of the polar trophectoderm differentiating to form the ectoplacental cone and the extraembryonic ectoderm, pushing the primitive ectoderm down into the blastocoel cavity in the process. The primitive endoderm around the epiblast differentiates into visceral and parietal endoderm which contribute to the yolk sacs. The visceral endoderm surrounds the egg cylinder and the parietal endoderm is found on the blastocoel face of the mural TE. Gastrulation begins at 6.5 dpc and by 10 dpc the basic body plan of the embryo is set up.

1.2.1. In vitro model systems of early mouse development.

It is not generally feasible to use whole embryos for studies on the control of transcription during early embryogenesis due to the small size of the blastocyst, the diversity of cell types and the asynchrony of early development. Two different tissue culture model systems are available which represent the pluripotential ICM cells, and which may be differentiated in vitro.

The older of the two systems uses cells derived from teratocarcinomas. These can arise spontaneously or be induced by grafting early embryos or primordial germ cells to ectopic sites. The teratocarcinoma OTT6050 was produced by grafting a 6 day embryo to the testis of a histocompatible host, and then established by serial transplantation through 30-40 generations (Stevens, 1970). Subsequent in vitro culture of pieces of OTT6050 gave rise to a number of embryonal carcinoma (EC) cell lines with different characteristics. These cell lines include F9 cells (Bernstine et al., 1973) and PCC4 and PCC3 cells (Nicolas et al., 1976), which all show undifferentiated morphology. PCC3 and PCC4 cells are defined as pluripotential - on transplantation to ectopic sites they give rise
to cells with EC cell morphology and differentiated cells characteristic of all three germ layers. Conversely F9 cells are defined as nullipotent, producing tumours with EC cell morphology only.

It is important to note that EC cells are transformed cell lines, and that part of their phenotype will reflect this transformation. Since the cell lines were set up they will have adapted to tissue culture, drifting still further away from the true ICM phenotype. These factors are reflected in the inability of EC cells to successfully repopulate the blastocyst during chimaera formation. EC cells can contribute to all the different tissues (Brinster, 1974; Papaioannou et al., 1975; Mintz and Illmensee, 1975; Martin, 1980; Stewart and Mintz, 1981) but the chimaeras formed often have abnormalities (Rossant and McBurney, 1982) and contribution to the germline is very inefficient. Despite these disadvantages EC cells remain a valuable experimental system, not least because of their ease of maintenance and manipulation in tissue culture.

The second, more recently developed system uses blastocyst cells placed directly in culture with no intermediate transformation step. The resulting cells are termed embryonic stem (ES) cells (Evans and Kaufman, 1981; Martin, 1981). Unlike EC cells, ES cells can efficiently contribute to all tissues of the blastocyst when used to generate injection or aggregation chimaeras with normal blastocysts (Bradley et al., 1984). While ES cells have the advantages of being non-transformed and truly totipotent, the difficulties of successfully maintaining these cells in culture mean that they are not always the model system of choice. To maintain the undifferentiated phenotype ES cells must either be grown on a layer of fibroblastic feeder cells, or supplied with differentiation-inhibiting-activity (Smith et al., 1988). It is also vital to plate out the cells within narrow limits of cell density and therefore they require frequent passaging.

It is possible to differentiate both EC and ES cells in vitro, thereby providing model systems mimicking differentiation of the ICM. F9 EC cells form primitive endoderm-like cells in the presence of all-trans retinoic acid. Further addition of dibutyryl cyclic AMP
induces formation of parietal endoderm-like cells (Strickland et al., 1980) or if put in suspension culture the primitive endoderm-like cells aggregate to form clumps surrounded by visceral endoderm-like cells (Hogan et al., 1981). It is also possible to derive cells resembling adrenergic neurones from F9 cells by treatment with nerve growth factor (Liesi et al., 1983). The different cell types are recognizable by their morphology and cell-surface antigens. The antibody anti-stage specific embryonic antigen 1 (anti-SSEA-1) is specific to undifferentiated cells (Solter and Knowles, 1981), whereas Troma-1 and Troma-3 are endoderm-specific (Boller and Kemler, 1983).

The advantage of the EC cell system is the ease with which differentiation is controlled; the necessary conditions are well defined and populations of a single cell-type are produced. It should be noted, however, that the parietal and visceral endoderm do not contribute to embryonic tissue, and that the underlying mechanisms controlling differentiation of primitive endoderm may differ from those used by primitive ectoderm. Studies on EC cells can nevertheless be valuable in demonstrating mechanisms by which cells can take differentiative decisions and, at least in the case of viral genes, EC cells and the early embryo seem to exhibit equivalent regulation of transcription.

ES cells once again provide a more accurate model of early embryogenesis. If placed in suspension culture small groups of cells will form "embryoid bodies" that morphologically resemble the normal egg cylinder (Evans and Kaufman, 1981). After about seven days of differentiation the "bodies" begin to develop cardiac muscles with intrinsic myogenic activity and a vascular system, showing the ES cells to be pluripotential. The disadvantages of this system are the asynchrony of differentiation and the variety of different cell types produced.

1.2.2. Gene expression during preimplantation development.

Little information is available concerning which genes are expressed during early mouse embryogenesis due to the lack of
LaRosa and Gudas (1988) showed that two proteins, one containing a homeobox, are encoded by the gene ERA-1, expression of which is induced by retinoic acid treatment of F9 EC cells.
It is known that zygotic transcription begins to take over from maternal transcription at the two cell stage (Flach et al., 1982), and that the need for enhancers to activate gene expression also appears at this time (Martinez-Salas et al., 1989). The presence of a number of proteins in preimplantation embryos has been investigated using enzymatic or immunological techniques (reviewed by Jackson, 1989). Few differences could be found with respect to the postimplantation embryo, suggesting that these methods are too insensitive to detect many of the proteins which are involved in the control of development.

Due to the difficulties of preparing sufficient embryonic material, detection of specific mRNAs during preimplantation development has relied heavily on the EC cell model system. Using this system a number of regulated genes have been identified. Wang and Gudas (1983) showed that the transcription of the type IV collagen and laminin genes is activated 20-fold by differentiation of F9 EC cells to parietal endoderm-like cells. Rosenthal et al. (1984) showed that the major histocompatibility gene complex (MHC) gene MHC H-2K^bm1 is activated upon differentiation of F9 EC cells to either parietal or visceral endoderm-like cells. Tilghman et al. (1985) showed that the a-fetoprotein gene is also activated on F9 EC cell differentiation. This activation correlates with activation during embryogenesis in the visceral endoderm of the developing yolk sac. Similarly, increased expression of the endo A gene on F9 EC cell differentiation correlates with activation in the developing endoderm of the blastocyst (Duprey et al., 1985). Decreases in expression during differentiation have also been observed, e.g. EC cells express constitutively high levels of the heat shock proteins which are down-regulated but inducible following differentiation (Mezger et al., 1989).

1.3 TRANSCRIPTION

Many of our ideas concerning the control of transcription in eukaryotes are based upon work with prokaryotes. Transcription has been widely studied in bacterial systems due to their relative
simplicity and ease of manipulation (reviewed by Beebee and Burke, 1988). *Escherichia coli* promoters have conserved sequences at -35 and -10 base-pairs (bp) from the start of transcription allowing recognition of the start site by RNA polymerase. RNA polymerase has been purified from several different strains of bacteria and shown to have a conserved pattern of subunits. The core polymerase consists of two α chains, one β chain and one β′ chain; to make a functional holoenzyme the sigma subunit is also required. There are multiple forms of the sigma subunit capable of conferring distinct properties on the RNA polymerase. Short DNA sequences in the vicinity of the promoter are bound by activator or repressor proteins to allow control of the initiation of transcription.

In contrast to prokaryotes, the eukaryotes have three different types of RNA polymerase. RNA polymerase II transcribes the protein coding (Class II) genes into heterogeneous nuclear RNA (hnRNA), the precursor of messenger RNA (mRNA). RNA polymerase II can be functionally distinguished from RNA polymerase I which transcribes ribosomal RNA, and RNA polymerase III which transcribes transfer RNA and other small RNAs by sensitivity to a bicyclic octapeptide, α-amanitin. The polymerase II enzyme is a complex of more than 500 kilo-daltons (kd), consisting of two large subunits of approximately 200kd and 140kd plus about ten smaller subunits. The two largest subunits have been cloned from several organisms and shown to have homology with subunits of *E. coli* RNA polymerase (Allison *et al*., 1985; Sweetser *et al*., 1987).

The correct start site for transcription is indicated by the presence of a promoter sequence. Many Class II promoters have a sequence known as the TATA box about 25 or 30 bp 5′ of the mRNA start site. Others (in particular the "house-keeping" genes) have a GC rich region upstream of the start; these promoters tend to have heterogeneous start sites. The development of *in vitro* transcription systems using cell-free extracts has enabled the identification and characterization of a number of general transcription factors, which are required in addition to RNA polymerase II for accurate and efficient initiation of transcription.

Early work by Matsui *et al.* (1980) identified four components
required for the active and selective initiation of transcription at the adenovirus 2 major late promoter (MLP) (nucleotides -56 to +33 with respect to the start of transcription). The four activities, transcription factors (TF) IIA, IIB, IIC, and IID were purified from a soluble extract of human KB cells by column chromatography. Later TFIIC was shown to be equivalent to poly (ADP-ribose) polymerase (Slatery et al., 1983). TFIIC suppresses random transcription in vitro by binding to template nicks and was found to be dispensable on further purification of the other factors. The factor TFIID is a DNA-binding protein, which interacts with the TATA box and start site of the MLP (Sawadogo and Roeder, 1985). Although the partially purified TFIID from HeLa cells (a transformed human cell line) has a molecular weight of approximately 100kd (Samuels et al., 1982) it has a functional equivalent of 27kd in the yeast Saccharomyces cerevisiae (Buratowski et al., 1988). Three separate groups have cloned the spt15 gene encoding the more stable yeast form of TFIID (Hahn et al., 1989a; Eisenmann et al., 1989; Horikoshi et al., 1989). Analyses of spt15 mutations have shown that the gene is essential for normal transcription in vivo (Eisenmann et al., 1989). Addition of in vitro translated TFIID to a mammalian cell extract that had been depleted of endogenous TFIID restored the ability to transcribe the adenovirus MLP (Hahn et al., 1989a), confirming the identity of the spt15 gene product.

More recently the genes encoding TFIID from HeLa cells (Peterson et al., 1990; Hoffman et al., 1990), Drosophila (Hoey et al., 1990) and Arabidopsis (Gasch et al., 1990) have also been isolated. Arabidopsis has two TFIID genes, but human, Drosophila and yeast have single TFIID genes which produce proteins with a highly conserved C-terminal region of approximately 180 amino-acids. The chromatographic fraction termed TFIID apparently contains components other than the cloned gene product which are involved in mediating the response to transcriptional activators (reviewed by Ptashne and Gann, 1990).

TFIID interacts not only with DNA, but also with the factor TFIIA; complexes of DNA and TFIID are only stable to electrophoresis through native gels in the presence of TFIIA (Hahn et al., 1989a). A yeast equivalent of mammalian TFIIA has also been identified, by
complementation of a depleted extract (Hahn et al., 1989b). Purified yeast TFIIA does not form specific complexes with DNA, but it does increase the region of interaction of TFIID with the TATA box by several bases upstream. Kinetic studies have shown that both TFIIA and TFIID are involved in the earliest steps of setting up the RNA polymerase II transcription complex (Hawley and Roeder, 1985; Reinberg et al., 1987), but the order of addition remains unclear. It has been proposed that TFIIA may alter the conformation of TFIID, or alternatively that TFIIA may act as a bridging molecule between TFIID and other constituents of the transcription complex. The recent cloning of TFIIA from HeLa cells (Zheng et al., 1990) will allow these models to be tested.

The binding of RNA polymerase II to the promoter requires a bound TFIID molecule (Reinberg et al., 1987), TFII B, E and F can then enter the complex. TFIIIE and F independently interact with purified RNA polymerase II, and are required for the formation of rapid start complexes. TFIIF has been purified to apparent homogeneity and exists as a heterodimer of 30 and 78kd polypeptides (Mermelstein et al., 1989). TFII B is a single polypeptide of 27kd (Mermelstein et al., 1989), which may be a bridging molecule connecting the TFIIA/D complex to RNA polymerase II and thus be involved in determining the distance from the TATA box to the start site.

Using RNA polymerase II affinity chromatography Sopta et al. (1985) isolated three RNA polymerase II associating proteins (RAPs) of 30, 38 and 74kd. RAP 38 is involved in elongation, but the 30 and 74kd proteins form a complex and are required for initiation of transcription (Burton et al., 1988). Antibodies to RAP 30 cross-react with the 30kd subunit of TFIIF (Flores et al., 1988) and both RAP 74 and the 78kd subunit of TFIIF are highly phosphorylated; these data strongly suggest that RAP 30 and 74 constitute TFIIF. Rap 30 has recently been cloned (Sopta et al., 1989) and found to have homology to the E. coli sigma factor (sigma is required for initiation of transcription in E. coli.) The RAP 30/74 complex has associated DNA helicase activity which presumably functions to melt the DNA at the promoter. This requires ATP, explaining the energy requirement of the initiation of transcription. The complete complex consisting of DNA,
TFIIA, TFIID, RNA polymerase II, TFIIB, TFIIE, TFIIF and ATP (with interactions occurring in roughly that order) is capable of initiation of transcription.

1.3.1. Control of initiation of transcription.

One of the best understood of all transcriptional control systems is the complex circuitry controlling the life cycle of phage lambda (reviewed by Ptashne, 1987). Lambda is a bacteriophage which on infection of a host cell can follow one of two alternative paths. Either the lytic cycle in which the phage rapidly multiplies and the host is lysed, or lysogeny in which the phage DNA integrates into the host genome to form a prophage. Most of the phage genes are not expressed during lysogeny; however, external stimuli such as UV light can induce expression and trigger the lytic cycle. The lambda repressor protein, cI, plays a vital role in the maintenance of lysogeny by binding to specific sites in the viral genome. As its name suggests the lambda repressor is a negative regulator of transcription, but it can also activate transcription from its own gene. The structural and functional properties of the cI protein are well understood. The following characteristics of cI are shared by some other prokaryotic and eukaryotic transcription factors:

i) cI binds to DNA via an α-helix - the "recognition" helix, which interacts with the major groove in the DNA double-helix. A second α-helix lies across the major groove to help orientate the recognition helix. This type of binding domain has become known as the "helix-turn-helix" motif.

ii) The DNA-binding and transcription control regions of cI are in two distinct domains. It has proved possible to interchange such domains and retain function in some cases (reviewed by Ptashne, 1989).

iii) The repressor binding site is palindromic and the repressor binds as a dimer; dimerization has also been observed in several other transcription factors.

iv) Two repressor binding sites occur in close proximity on the lambda genome and binding shows cooperativity, this allows increased
sensitivity to small changes in protein concentration.

v) If the repressor binding sites are moved apart on the viral genome the repressor protein molecules continue to show cooperativity of binding, the intervening DNA looping out to allow the molecules to interact. This observation has led to the hypothesis that a "looping" mechanism may also explain how eukaryotic transcription factors act over large distances.

The large and detailed body of knowledge that has been accumulated concerning the lambda phage transcriptional control circuits, and other prokaryotic systems such as the lac and trp operons, has helped to direct experiments on eukaryotic systems, and has aided in the interpretation of the results of such experiments. It is interesting that in spite of the considerable differences between the transcription machinery of prokaryotes and eukaryotes, many basic concepts are common to both.

A unique feature of eukaryotic DNA is that the majority is tightly packaged to form chromatin. Every 200 bp of eukaryotic DNA is associated with histone proteins to form repeating structures termed nucleosomes. The nucleosomes are further wound into higher order structures which condense the DNA to a thousandth of its original length (reviewed by Lewin, 1980). Before transcription can occur the DNA is partially decondensed to allow interactions with regulatory proteins. Nucleosome free regions have been observed in GC-rich promoters (Tazi and Bird, 1990). The mechanisms controlling decondensation are not fully understood, but this step offers a potential point for regulation not available in prokaryotic organisms.

The general transcription factors described previously are required for accurate initiation of transcription at all Class II promoters. The rate of initiation is controlled by the interaction of specific transcription factors with particular cis-acting sequence elements, both 5' and, more rarely, 3' of the transcription start site. Mutation of the control sequences has shown that every gene has its own unique array of transcription factor binding sites mediating both activation and repression of transcription. The control sequences have been functionally divided into two broad classes termed
promoters and enhancers. The region immediately 5' of the start site is termed the promoter and is required for accurate initiation of transcription. The activity of a promoter sequence is dependent on its position and orientation with respect to the transcriptional start. Many promoters incorporate a TATA box or GC rich regions as discussed above. Other common promoter elements are the GC box (GGCGGG, the consensus binding site for transcription factor Spl), and the CCAAT box, both of which can function in either orientation and tend to be found 40 to 100 bp upstream of the transcriptional start.

Enhancers are defined by their ability to act in an orientation and distance independent fashion to control the rate of initiation of transcription. Enhancer elements occur both 5' and 3' of the transcription start site, and in some cases are active from as far away as 30 kilobases (kb). The definitions of promoter and enhancer can only be made at the functional level as particular transcription factor binding sites may occur in either situation. Another common property of enhancers is a modular structure. This is especially well understood in the case of the SV40 enhancer, which incorporates several, functionally interchangeable sequence modules (reviewed by Jones et al., 1988). It has also been discovered that particular sequence modules may occur in more than one type of enhancer. The different combinations of transcription factor binding sites found in the control region of each gene facilitates complex transcriptional control. Genes can be activated and inactivated in response to a wide range of stimuli including hormones, growth factors, temperature changes, presence of metal ions and cell cycling. Similarly, transcription can be strictly limited to a particular cell type, region of the body or stage of development.

The mechanism by which transcription initiation is activated or repressed by factors binding thousands of bases away from the transcription start site remains unclear. Various models have been proposed, the main ones are "looping" which assumes the intervening DNA sequences loop out to allow transcription factors to interact directly with the transcription apparatus, and "sliding" which assumes that once a protein has recognized its consensus binding site
The Cx proteins include an important class of ligand activated transcription factors, the steroid hormone receptors (reviewed by Beato 1989).
it then slides along the DNA until it reaches the transcription complex (reviewed by Ptashne, 1986). The looping model is perhaps the more widely supported, but it remains to be conclusively proved.

1.3.2. Transcription factors.

Many different transcription factor proteins have been purified and several have also been cloned. Analysis of the structure and function of these proteins has revealed that they have separate DNA-binding and transcription activation domains which frequently fall into one of a few broad classes. The most common types of DNA-binding domain are zinc-finger, helix-turn-helix or basic motifs, although transcription factors have been characterized which fall into none of these groups. Zinc-finger motifs were first recognized in the RNA polymerase III transcription factor TFIIIA (Miller et al., 1985). Since then two different types of zinc-finger motif have been found to facilitate DNA-binding in several different transcription factors. The TFIIIA-like, or C\textsubscript{2}H\textsubscript{2}, zinc-fingers are domains in which a Zn\textsuperscript{2+} ion is co-ordinated by a pair of cysteine and a pair of histidine residues divided by a loop of 12 amino acids. The C\textsubscript{x} proteins have a variable number of cysteine residues available for chelation (reviewed by Evans and Hollenberg, 1988).

The helix-turn-helix domain was identified as a DNA-binding domain in prokaryotic repressor proteins as discussed above, and in the specialized form of the homeodomain in several developmentally regulated Drosophila melanogaster transcription factors. Similar domains have more recently been found in mammalian octamer transcription factors (reviewed by Schaffner, 1989). The third class includes a group of transcription factors that require homo- or hetero-dimerization for efficient DNA-binding; dimerization is mediated by a structure termed the "leucine zipper" in which leucine residues occurring every seventh residue allow hydrophobic interactions between two subunits. A basic region adjacent to the leucine zipper motif in each subunit is required for DNA-binding (reviewed by Mitchell and Tjian, 1989).

Transcriptional activation domains also fall into three broad
classes - acidic, glutamine-rich and proline-rich (reviewed by Mitchell and Tjian, 1989). Acidic domains are characterized by negative charge and the potential to form α-helices. Despite negligible primary sequence homology such domains have proved functionally interchangeable between yeast transcription factors (reviewed by Ptashne, 1989). Similarly, the glutamine-rich domain of the *D. melanogaster* protein *Antennapedia* can partially substitute for the glutamine-rich activation domain of Spl (Courey and Tjian, 1988). Glutamine-rich domains have also been observed in several other transcription factors. A proline-rich domain has been identified as the transcriptional activation domain in the CCAAT box transcription factor, CTF (Mermod et al., 1989). Once again several other transcription factors have proline-rich domains which may prove to mediate activation.

The current high level of interest in the control of gene expression is leading to the identification of many new transcription factors. Although a list of more than 120 transcription regulating proteins was compiled by Wingender (1988), this is now far from complete. Some of the most studied and best understood examples are discussed below.

### 1.3.3. Transcription factor Spl.

Spl is one of the best characterized mammalian transcription factors. It was first identified as a transcription factor present in human (HeLa) cell extracts, capable of specifically activating transcription from the SV40 early promoter *in vitro* (Dynan and Tjian, 1983a). It binds to the 21 bp repeat sequences upstream of the SV40 early promoter which contain six copies of the consensus GGGCGGG (Dynan and Tjian, 1983b). Deletion of the 21 bp repeats reduces expression by 10-fold in transient transfection experiments (Everett et al., 1983). Spl also activates transcription from a variety of other viral and cellular promoters which contain the GC-box motif or closely related sequences. These include the herpes simplex virus (HSV) thymidine kinase (tk) gene, the HSV immediate-early 3 gene, the human immunodeficiency virus LTR, the human metallothionein II A gene,
and the mouse dihydrofolate reductase gene (reviewed by Kadonaga et al., 1986). The number of Spl binding sites varies between one and six in the above examples; a single site allows activation of transcription and in situations where there are multiple binding sites the one closest to the transcriptional start seems to be the most important. Despite the asymmetry of the binding site it is functional in either orientation. Binding sites frequently occur in close proximity to other transcription factor binding sites and it is likely that these sites act synergistically.

There are about 5000-10000 Spl molecules in each HeLa cell. The protein has been purified to homogeneity by oligonucleotide affinity chromatography (Kadonaga and Tjian, 1986; Briggs et al., 1986). Two major forms of the protein exist of 105kd and 95kd as determined by SDS-PAGE analysis. The higher molecular weight form is phosphorylated at several serine residues and can be converted to the lower molecular weight form by phosphatase treatment (Jackson et al., 1990). Availability of purified protein allowed a partial amino acid sequence to be determined and from this oligonucleotides containing the coding sequence were designed. These oligonucleotides were used to isolate cDNA clones encoding the C-terminal 80% of Spl from a HeLa cell cDNA library (Kadonaga et al., 1987; 1988). Sequence inspection shows that the C-terminus of Spl has three potential zinc-finger domains of the TFIIIA-like class. DNA-binding and transcription activation have been shown to require the presence of zinc ions suggesting that the zinc-finger domains are functional (Kadonaga et al., 1987; Westin and Schaffner, 1988).

The functional domains of Spl have been analyzed in detail by the expression of wild-type and mutant forms of the protein in a D. melanogaster cell line (Schneider line 2), which itself has no Spl or related activity (Courey and Tjian, 1988). Using this approach two glutamine-rich transcription activation domains have been identified in the N-terminal half of Spl; these can be functionally substituted by the Antennapedia glutamine-rich domain (Courey et al., 1989). The same domains were identified by in vitro transcription assays using bacterially synthesized mutant forms of Spl (Kadonaga et al., 1988). Studies using Schneider cells have also shown that two Spl sites, one
Close inspection of SDS-PAGE analysis of Spl reveals that each major form comprises several sub-species, this reflects a further modification of the protein. Spl and several other RNA polymerase II transcription factors bear terminal N-acetyl glucosamine residues (GlcNAc; Jackson and Tjian, 1988). In the case of Spl the GlcNAc is in the form of monosaccharides attached via O-glycosidic linkages. Each molecule bears approximately eight GlcNAc residues. The glycosylated HeLa cell Spl is 3 to 5 times more active than unglycosylated E. coli synthesized Spl in vitro. Glycosylation does not affect the efficiency of DNA binding, suggesting that glycosylation plays a role in transcriptional activation, either directly or by facilitating protein-protein interactions (Jackson and Tjian, 1988). It is possible that glycosylation is involved in a common mechanism of transcription activation.

Experiments using cloned TFIID suggest that a further component of the TFIID chromatographic fraction is required for the activation of transcription by Spl (Pugh and Tjian, 1990). This type of additional protein has been termed a "coactivator" and may facilitate complex formation between Spl and TFIID or alternatively it may be required for a step in activation that occurs after formation of an Spl-TFIID complex.

1.3.4. The CCAAT box-binding family.

The CCAAT box sequence shares several properties with the Spl binding site - it occurs in multiple promoters, is in a similar position (40-100 bp upstream of the transcriptional start site), can function in either orientation, and is often found close to other transcription factor binding sites. This single sequence is recognized by a "family" of diverse proteins (several other examples exist of families of transcription factors).

The CCAAT box was first recognized not as a transcription factor binding site but as a binding site for NF-1 (nuclear factor 1), which directs initiation of adenovirus DNA synthesis in vitro (Nagata et
a7., 1982). The complete consensus sequence for NF-1 is TGGCT(N)₃AGCCAA. Mutational analysis of the HSV tk gene promoter (Jones et al., 1985) revealed that CCAAT-binding transcription factor (CTF) interacts with the CCAAT motif. Optimal expression from the tk promoter requires the CCAAT box and the two Spl binding sites which flank it. Purification of CTF revealed a family of proteins of 52-66kd (Jones et al., 1987) which corresponded to NF-1 at both the structural and functional level. Members of this protein family can stimulate both transcription and DNA replication, and are now referred to as CTF/NF-1 proteins. Multiple cDNA species encoding different members of the family were isolated by Santoro et al. (1988). The different CCAAT-binding proteins are probably products of differential RNA splicing, and may also have varying post-translational modifications. Some forms of CTF/NF-1 are O-glycosylated (Jackson and Tjian, 1988). Different murine cell lines apparently possess varying concentrations and forms of CTF/NF-1 (Goyal et al., 1990) which may confer tissue-specific control of gene expression.

The functional domains of one species of CTF/NF-1 protein were analyzed by expressing mutant forms in E. coli and D. melanogaster cells, using a similar approach to that described for Spl (Mermod et al., 1989). A region of about 200 amino acids at the N-terminus of the protein is responsible for DNA-binding and dimerization. The region is α-helix permissive but shows no homology to previously characterized DNA-binding domains, and it remains to be determined whether dimerization is a pre-requisite for DNA-binding. This same region is sufficient to activate initiation of adenovirus replication in vitro. Activation of transcription requires the proline-rich C-terminal 100 amino acids; this region remains functional if fused to the Spl zinc-finger DNA-binding domain.

CTF/NF-1 is not the only transcription factor which recognizes the CCAAT box homology. Graves et al. (1986) identified a rat liver CCAAT binding protein (CBP) which interacts with the CCAAT homologies of the HSV tk promoter and the Moloney murine sarcoma virus (Mo-MSV) LTR. Johnson et al. (1987) used a DNase 1 footprinting assay to identify a 20kd rat liver protein, the enhancer binding protein 20
(EBP 20) which interacts with the core enhancer homology
(5'TGTGCA^A^A^A^A^G 3') of Mo-MSV, SV40 and polyoma. This protein proved
functionally indistinguishable from CBP. A single cDNA has been
isolated which encodes a 42kd protein capable of binding both CCAAT
and enhancer core homologies (Landschulz et al., 1988). The
previously purified 20kd species is apparently a partially degraded
form of the protein which is now termed C/EBP.

The DNA-binding domain of C/EBP is found at the C-terminus of the
protein (Landschulz et al., 1989). This region includes a "leucine
zipper" domain responsible for dimerization, and a basic DNA-binding
domain. The leucine zipper has a leucine at every seventh residue and
the potential to form an α-helix. It has been hypothesized that the
α-helices from two C/EBP molecules lie anti-parallel to one another
in a coiled-coil, held together by hydrophobic interactions to create
a bipartite basic DNA-binding domain. This model is supported by the
dependence of DNA-binding on dimerization.

C/EBP expression is not limited to the liver - it is also detected
in gut, lung, adipose tissue, adrenal gland and placenta of the mouse
(Birkenmeier et al., 1989). Expression is apparently confined to
terminal differentiated cell types, e.g. onset of expression
correlates with differentiation of preadipocytes (Christy et al.,
1989). C/EBP has also been shown to activate the transcription of
cellular genes, including two adipocyte-specific genes (Christy et
al., 1989), and the serum albumin gene in cultured hepatoma cells
(Friedman et al., 1989). It is possible that the C/EBP gene product
plays a central role in establishing and maintaining the
differentiated, nonproliferative state.

Raymondjean et al. (1988) identified the albumin CCAAT factor
(ACF), which binds to the rat albumin promoter. It is distinct from
CTF/NF-1 and C/EBP and is present in liver and spleen. Chodosh et
al. (1988a) have identified two further HeLa cell CCAAT box binding
proteins, CPI and CP2. Each of these consists of multiple
heterologous subunits and has slightly different DNA-binding site
sequence constraints. CPI has been shown to consist of CPI A and B,
which are closely related to the yeast genes HAP2 and HAP3. The
subunits are functionally interchangeable between the yeast and human
forms (Chodosh et al., 1988b). Two further distinct CCAAT binding proteins have been identified by Dorn et al. (1987):- (i) NF-Y which recognizes the Major Histocompatibility Complex (MHC) Class II Y element CCAAT box, and (ii) NF-Y* which interacts with the HSV tk promoter. The functional significance of these multiple factors remains to be investigated.

1.3.5. Negative regulation of transcription.

The eukaryotic transcription factors described so far have been positive regulators of transcription, but recently several negative regulators have also been identified. In earlier work experiments were aimed at finding cis-acting sequences which activate transcription, thereby selecting for positive regulators.

There are a number of potential mechanisms by which transcription may be negatively regulated. The following models have been suggested:-

i) Repression involves competition between activator and repressor proteins for the same or closely adjacent binding sites.

ii) "Quenching", in which interactions between an activator and a repressor prevent the activator from performing its usual function; in this model the repressor protein may or may not bind to the DNA itself.

iii) "Squelching", in which the over-expression of an activating protein allows it to sequester other factors in solution preventing their interaction with the promoter.

iv) The repressor interacts directly with the initiation complex to prevent transcription.

Some systems in which negative regulation of transcription has been described are considered below.

One relatively well characterized negative regulator, IκB, seems to act via a form of the "quenching" model, in which it prevents migration to the nucleus of an activator of transcription. The positive transcription factor NF-κB has binding sites on the κ light chain enhancer and the HIV enhancer. The protein is only normally present in cell lines derived from mature B cells but can be induced
in other cell types by the phorbol ester TPA (12-O-tetradecanoyl-
phorbol 13-acetate; Sen and Baltimore, 1986a; 1986b). An inactive,
cytosolic form of NF-κB can be activated by treatment with the
dissociating agent sodium deoxycholate. A cytosolic inhibitor
protein, IkB, complexes with NF-κB and prevents entry into the
nucleus in pre-B cells (Baeuerle and Baltimore, 1988). IkB is
dissociated from NF-κB by sodium deoxycholate and also on TPA
treatment or B cell maturation. TPA stimulates protein kinase C to
phosphorylate IkB rendering it incapable of complexing with NF-κB
(Ghosh and Baltimore, 1990). Such a phosphorylation event is probably
also capable of disrupting the complex, effectively activating NF-κB.

Several DNA-binding proteins have been identified that include a
helix-loop-helix (HLH) domain and a basic DNA-binding domain, these
include the MyoD protein which is involved in muscle determination
(Davis et al., 1987). Benezra et al. (1990) isolated a new member of
this family with the potential to negatively regulate other members.
The clone isolated has a gene product with an HLH domain which
facilitates specific association with other HLH proteins, including
MyoD. However, the protein does not have a basic DNA-binding domain
and the heterodimers are unable to bind DNA. The clone has been named
Id for "inhibitor of DNA-binding". Transfection experiments indicated
that overexpression of Id inhibits transactivation of the muscle
creatinine kinase enhancer by MyoD.

One much studied but little understood repressor of transcription
is the product of the adenovirus early gene E1A, which can act as
both an activator and a repressor of transcription. The E1A gene
product is required for expression of the other adenovirus early
genes and can also activate some cellular genes (reviewed by Berk,
1986). However, E1A can also repress enhancer mediated transcription.
In co-transfection experiments the SV40 and polyoma enhancers are
inhibited, together with the E1A enhancer itself (Borrelli et al.,
1984; Velcich and Ziff, 1985). Similarly the immunoglobulin heavy and
k light gene enhancers are inhibited by E1A in transfection and
infection experiments (Hen et al., 1985). Repression of transcription
from muscle-specific promoters and the insulin gene promoter has also
been reported (Stein and Ziff, 1987; Webster et al., 1988). E1A does
not have a specific DNA-binding site and therefore must both activate and repress transcription using indirect mechanisms. It has recently been shown that each separate module of the SV40 enhancer is individually subject to repression by E1A (Rochette-Egly et al., 1990). E1A also represses transcription via other transcription factor binding sites, including the AP1 sites of the polyoma enhancer and the H2TF-1 site of the MHC class 1 H-2K\(^b\) gene. The N-terminal domain of E1A is apparently crucial for repression but the mechanism remains unresolved. It has been proposed that an intermediary factor may transmit the effects of E1A to target transcription factors, or alternatively that E1A may inactivate other factors by facilitating post-transcriptional modifications.

The product of the proto-oncogene c-fos can also behave as both an activator and a repressor of transcription. The c-fos protein complexes with members of the jun protein family via a leucine zipper domain to produce the AP-1 factor, an activator of transcription. AP-1 sites have been recognized in the human metallothionein II\(\alpha\) gene promoter and in multiple TPA inducible genes. Transcription of the heat shock protein (hsp) 70 gene and of c-fos itself is down-regulated by the c-fos protein. Lucibello et al. (1989) have shown that the autoregulation is not mediated by the AP-1 site in the c-fos promoter, but by the serum response element (SRE). Although the DNA-binding domain is not necessary for repression, the leucine zipper domain is necessary, suggesting that an interaction with a second protein is required.

Collagenase is an example of a TPA inducible gene in which induction is mediated via an AP-1 binding site. This same site also mediates repression by glucocorticoids due to inactivation of AP-1 by the glucocorticoid receptor (GR). Several groups have recently shown that c-jun and GR mutually repress one another, presumably by forming a stable but inactive heterodimer (Jonat et al., 1990; Schule et al., 1990; Yang-Yen et al., 1990).

Transient expression assays have shown that transcription of the D. melanogaster gene engrafted is positively regulated by the other homeobox proteins fushi tarazu, paired and zen. However, the homeobox proteins even-skipped and engrafted itself cause repression of
engrailed transcription (Han et al., 1989). The 96 bp promoter fragment used in these experiments contains five copies of a 10 bp sequence motif which is recognized by each of the homeobox proteins tested. This result implies that competition may occur between different transcription factors for the same cis-acting sequence element. Small variations in concentration of particular homeodomain proteins could therefore have profound effects on the rate of transcription of target genes.

The examples of negative regulators of transcription discussed above illustrate the diversity of potential repression mechanisms. Now that the prevalence of repressors of transcription has been appreciated it is likely that more examples will be reported in the future.

1.4 VIRAL TRANSCRIPTION

Many of the transcription factors previously discussed were first identified as regulators of viral transcription; viral genes tended to be used in early work as they were already cloned and sequenced. Most of these factors were later found to also act on cellular genes, therefore experiments on the control of viral gene expression were a useful starting point in the study of control of transcription. Fortuitously viral enhancers proved to be modular in structure allowing the identification of multiple transcription factor binding sites. Selected aspects of this large field are discussed below.

1.4.1. Simian virus 40.

SV40 is a small DNA virus belonging to the papovavirus family. The SV40 enhancer was the first to be described (Banerji et al., 1981); it remains the best characterized of all enhancers and is often considered a prototype. The enhancer lies between the early and late promoters of SV40 and activates transcription from both. It can also activate a wide variety of heterologous promoters and act in an orientation and distance independent fashion (reviewed by Hatzopoulos et al., 1988). These properties have become the criteria by which an
enhancer is defined. Deletions and point mutations have been generated across the entire enhancer, providing a detailed analysis of its sequence requirements (Zenke et al., 1986). From these studies it is known that the enhancer falls into three main domains A, B and C. Although both domains A and B are required for full activity, they are functionally redundant, i.e. if one is deleted its activity can be replaced by a duplication of the other. The activity of domain C only becomes apparent following mutation of the other two domains. Multimers of any one of the three domains can act autonomously as an enhancer element (Ondek et al., 1987).

A number of transcription factor binding sites have been identified within the enhancer, some of which also occur in other enhancer sequences (reviewed by Jones et al., 1988). These include the enhancer core homology, which has been reported to interact with AP-2 and AP-3. It is possible that AP-3 is equivalent to C/EBP. The AP-2 transcription factor also acts on the control sequences of many other genes including human metallothionein II\textsubscript{A}, collagenase and proenkephalin. Mitchell et al. (1987) noted cooperation between AP-2 and Spl, but an inhibition of AP-2 binding by the SV40 large T-antigen. The factor AP-4 recognizes a site in domain A and acts synergistically with AP-1 to activate transcription (Mermod et al., 1988). The SV40 enhancer also has a degenerate "octamer" binding site (consensus ATGCAAAT). The octamer sequence has been identified in many different genes - it is the recognition site for a family of octamer transcription factors some of which confer tissue-specific expression (Scholer et al., 1989). A variety of proteins with different cell specificities have been reported to interact with this motif in the SV40 enhancer (Rosales et al., 1987). Several other transcription factors have also been shown to interact with the SV40 enhancer (reviewed by Jones et al., 1988; La Thangue and Rigby, 1988b).

1.4.2. Polyoma virus.

Polyoma is another member of the papovavirus family, of a similar size and genome organisation to SV40. The polyoma enhancer has been
localized to a 246 bp fragment which is required for both DNA replication and expression from the early promoter (Tyndall et al., 1981). Mutation analysis of the region has allowed identification of two separate elements each of which can function independently as an enhancer (Veldman et al., 1985). The A enhancer has greater activity than the B enhancer. Veldman et al. (1985) further subdivided the region into domains A, B, C and D by functional genetic analysis.

The 22 bp A domain (nucleotides 5108-5130) has intrinsic enhancer activity if multimerized. Piette and Yaniv (1987) have shown that two nuclear proteins from mouse 3T6 cells interact independently with this domain. The factors are termed PEA-1 and 2 for polyoma enhancer A binding factors. PEA-1 is probably the murine equivalent of the human transcription factor AP-1 (Lee et al., 1987). Like AP-1, PEA-1 activity is TPA inducible (Wasylyk et al., 1987). Another polyomavirus enhancer-binding protein, PEBP5, with a binding site overlapping that of PEA-1 has recently been described (Asano et al., 1990). PEBP5 is also responsive to TPA. Domain A has also been reported to interact with a fourth factor, AP-4 (Jones et al., 1988).

Ostapchuk et al. (1986) described a factor, EF.C which binds to domain C. Footprinting assays show that EF.C interacts with a region almost precisely equivalent to that identified functionally by Veldman et al. (1985). Piette and Yaniv (1986) have described a factor PEB1 (polyoma enhancer binding factor 1) which interacts with the B enhancer. The binding site of PEB1 is homologous to the SV40 sequence which interacts with AP-2 and -3. It is possible that the PEB1 factor comprises AP-2 and -3.

1.4.3. Adenovirus.

The adenoviruses are non-enveloped icosahedral viruses with a linear double-stranded DNA genome of 30-35 Kbp. Their organisation is more complex than that of the papovaviruses reflecting the larger genome. The first gene to be expressed following infection is the early gene E1A, which encodes a transcription factor capable of stimulating transcription from all the early genes including its own (Nevins, 1981). The E1A gene produces two RNA transcripts early in
infection, the 12S and 13S RNAs, which are translated to produce proteins of 243 amino-acids and 289 amino-acids respectively. The two gene products have multiple activities. They are activators of transcription from viral and endogenous genes, including heat shock protein 70 (hsp70), β-tubulin, and proto-oncogenes (Berk et al., 1979; Kao and Nevins, 1983; Sassone-Corsi and Borrelli, 1987). They can also repress transcription as discussed above, and in cooperation with E1B or other oncogenes can transform cells.

Three conserved domains have been identified in the E1A protein by sequence comparison between different strains of adenovirus. Transcriptional activation is strongly dependent upon E1A region 3 (amino-acids 140-188) which is unique to the 289 amino-acid form of the protein (Lillie et al., 1987; Fahnestock and Lewis, 1989). A synthetic peptide (PD3) corresponding to 49 amino-acids of region 3 plus the first three amino-acids of the second exon can autonomously activate transcription from the E2 promoter (Lillie et al., 1987). The inhibition of cellular protein synthesis does not prevent activation of transcription by PD3 (Green et al., 1988), implying that E1A does not activate transcription by inducing increased expression of other transcription factors. PD3 has the potential to form a zinc-finger structure, of the type previously implicated in DNA-binding. However no sequence-specific DNA-binding activity has been demonstrated for E1A.

Lillie and Green (1989) showed that fusions of E1A region 3 and the yeast transcription factor GAL4 DNA-binding domain can activate transcription from promoters containing binding sites for GAL4. E1A region 3 is acidic and can be functionally replaced within the context of the fusion by the activating region of an HSV transcription factor, VP16. In the absence of GAL4 binding sites the fusion can still activate transcription provided that the C-terminus of region 3 remains intact. These results suggest that E1A has a transcription activation domain which acts at the promoter in a similar fashion to more common DNA-binding transcription factors, and that it is directed to the promoter by the C-terminus of region 3. It is, however, possible that the normal mechanism of E1A transcription activation is perturbed by use of a fusion protein.
As E1A is not a DNA-binding protein it must act indirectly, possibly using protein-protein interactions to mediate interactions with the transcription machinery, as is suggested by the results of Lillie and Green (1989), or alternatively by modifying other transcription factors. One likely candidate for E1A modification is the activating transcription factor (ATF), which has binding sites in the promoters of several E1A inducible genes (Lee et al., 1987).

Another candidate is E2F which has binding sites in both the E2 promoter and the E1A promoter (Kovesdi et al., 1987a). The activity of this activator of transcription is apparently increased by adenovirus infection in an E1A dependent manner (Kovesdi et al., 1987a; Hardy et al., 1989). The product of another early gene, E4, is also involved in increasing the activity of E2F (Babiss, 1989; Hardy et al., 1989; Huang and Hearing, 1989). Jansen-Durr et al. (1989) reported that E1A does not affect the DNA-binding activity of E2F, but rather favours simultaneous binding of two E2F molecules to the two recognition sites in the E2A promoter. They hypothesized that E1A activates transcription by facilitating cooperative dimerization. Hardy et al. (1989) used a gel retardation assay to show that several protein-DNA complexes form on the E2 promoter, and that these complexes are altered following infection. They found that presence of the infection-specific complex is dependent upon the E4 gene product, but that its activity is E1A dependent. The results of Neill et al. (1990) also showed that E1A and E4 gene products both play a role in the activation of E2F mediated E2 gene transcription. E2F is probably a family of transcription factors which vary according to cell type and are modified or regulated in response to adenovirus infection.

Simon et al. (1988) reported that only certain precise sequences at the TATA element, including that of the hsp70 promoter but excluding that of the SV40 promoter, are capable of mediating E1A activation. However, these results are contradicted by two other groups. Williams et al. (1989) used transfection assays to show that any mutation of the human hsp70 promoter which affects basal levels of transcription also affects E1A mediated activation of transcription. Taylor and Kingston (1990) considered the ability of E1A to activate a variety
of nonsense and substitution mutants of the human hsp70 promoter in transfection and infection experiments. They too found no clear correlation between activation and a particular factor binding site. They did however note an inverse correlation between strength of promoter and degree of E1A activation. Pei and Berk (1989) used synthetic promoter sequences to analyse which transcription factor binding sites could mediate activation of transcription by E1A. The ATF binding site, the TATA elements of E4 and E1B and two E2F binding sites in the configuration found in the E2 promoter facilitated activation, but a single E2F site or an Sp1 site did not. Considered as a whole the literature suggests that E1A can indirectly activate transcription via numerous sequence elements. A similar conclusion has been reached concerning E1A mediated repression of transcription (Rochette-Egly et al., 1990). The details of the mechanisms by which the E1A protein acts to modify the rate of initiation of transcription await clarification.

1.4.4. Viral transcription in embryonic stem cells.

Experiments using viruses have proved particularly useful in characterizing the transcriptional regulation of embryonic stem cells. This work has been concentrated upon embryonal carcinoma (EC) stem cells due to their relative ease of manipulation (discussed in section 1.2.1.). The vast body of knowledge introduced above, concerning the control of viral transcription in differentiated cell types, has allowed comparisons to be drawn between stem cells and their differentiated derivatives. These types of analyses have shown that stem cells possess a novel array of transcription factors. The activators and repressors of transcription present in a particular type of stem cell presumably play an important role in maintaining the stem cell phenotype. Conversely, changes in transcriptional control must accompany differentiation and development.

The papovaviruses and the C-type retroviruses are unable to replicate in EC cells, due to a block to the transcription of the early genes (Swartzendruber and Lehman, 1975; Swartzendruber et al., 1977; Teich et al., 1977). The enhancers of these viruses are
apparently inactive in EC cells, which could be due to the absence of positive transcription factors or the presence of negative transcription factors. The block to retrovirus replication is discussed in section 1.5.4. A number of polyoma viruses with mutations in the enhancer region have been isolated which are capable of expression in EC cells. All mutants capable of expression in the PCC4 cell line have undergone extensive rearrangement, usually involving duplication of the A domain and deletion of the B domain (Katinka et al., 1981; Melin et al., 1985). Temporary inhibition of protein synthesis can alleviate the block to viral replication in PCC4 cells (Cremisi and Babinet, 1986). This suggests that PCC4 cells contain a labile inhibitor of transcription which prevents the initiation of early gene transcription.

Mutant viruses able to replicate in the F9 EC cell line generally have point mutations in the enhancer. Fujimura et al. (1981) described three mutants capable of replicating in F9 cells, Py F101, 111 and 441 each of which has a point mutation at nucleotide 5233 in the B enhancer. Mutants F101 and 111 also have tandem duplications encompassing this region. Kovesdi et al. (1987b) used oligonucleotides corresponding to polyoma wild-type or F9.1 (equivalent to F441) mutant sequences in a gel retardation assay, to identify a protein present in both F9 and fibroblast cells capable of interacting only with the mutant sequence. The same site was shown to interact with purified CTF/NF-1 protein (Tseng et al., 1988). Another member of the CCAAT box binding protein family could be acting at this site in vivo, particularly as it has been suggested that CTF/NF-1 protein is present only at low levels in EC cells (Speck and Baltimore, 1987).

Sassone-Corsi et al. (1987) used co-transfection assays to show that an excess of the wild-type polyoma virus regulatory region can titrate out a negative factor repressing transcription in F9 cells. The F9.1 mutant sequence is unable to perform this function, suggesting that the mutation has disrupted the binding site of a negative regulator of transcription. It thus remains unclear if the mutation in polyomavirus F9.1 is creating the binding site for a positively acting transcription factor, or destroying the binding
site for a negative regulator of transcription. Kryszke et al. (1987) report that F9 cells have almost undetectable levels of PEA-1, a positive regulator of polyomavirus transcription, unlike their differentiated derivatives. Similarly, the level of PEBP5 also increases following EC cell differentiation (Asano et al., 1990).

Lack of PEA-1 or PEBP5 could be partly responsible for the inactivity of the polyoma enhancer in EC cells. Transfection experiments in F9 and fibroblast cells support the hypothesis that F9 cells have low levels of PEA-1 (Wasylyk et al., 1988). However this group also suggest that PEA-2, generally considered to be an activator of transcription, is in fact a repressor which negatively regulates transcription in EC cells. Their experiments make use of multimerized synthetic binding sites acting upon a heterologous reporter gene, and the results obtained may not be valid in the context of the viral enhancer. If PEA-2 is a negative regulator of transcription its effects can obviously be overridden by mutations at other binding sites in the enhancer.

The SV40 enhancer is also inactive in EC cells (Gorman et al., 1985b). This inactivity is however less well understood, because no mutant forms of the virus capable of replicating in EC cells can be isolated. Co-transfection experiments have shown that the polyomavirus enhancer region can compete for a negative regulator of transcription to allow a low level of SV40 enhancer activity (Sleigh et al., 1987). This suggests that the mechanisms determining inactivity of the two viruses in EC cells share common properties and that a repressor of transcription is involved. Further experiments by Sleigh (1987) used cycloheximide, an inhibitor of protein synthesis, to show that a labile repressor of transcription is active in both differentiated and undifferentiated cells. La Thangue and Rigby (1988a) have mimicked the SV40 enhancer regulation in vitro in whole-cell extracts of both F9 EC cells and their differentiated derivatives. Experiments in which the two types of extract are mixed suggest that the differentiated cell type is dominant. This result supports the hypothesis that EC cells are deficient in activators of transcription. As previously discussed, PEA-1 (the murine equivalent of human API, a factor required for SV40 transcription) is found at
virtually undetectable levels in EC cells. The paucity of PEA-1 and other transcriptional activators may contribute to the inactivity of the SV40 enhancer in EC cells. It seems likely that both of the potential causes of enhancer inactivity, i.e. absence of positive transcription factors and presence of negative factors, play a role in the inactivation of papovaviruses in EC cells.

Imperiale et al. (1984) showed that adenoviruses, unlike papovaviruses, are able to replicate normally in EC cells. Surprisingly they also found that the d7312 virus, an E1A mutant which permits only low level expression of the early gene E2 in HeLa cells, allows wild-type E2 expression levels in EC cells. They further noted that the E1A inducible gene hsp70 is also expressed at high levels in EC cells. These results suggest that EC cells possess an E1A-like activity, able to activate transcription in an analogous fashion to the E1A protein itself. Nevin's group have reported that the E2F factor is involved in adenovirus E2 transcription, and that its activity is increased by E1A (Kovesdi et al., 1987a). Using exonuclease III protection and gel retardation assays they have shown that a similar factor is present at high levels in F9 EC cells in the absence of E1A, but not in their differentiated derivatives (Reichel et al., 1987).

La Thangue and Rigby (1987) performed in vitro transcription reactions and found that E1A inducible promoters were more active in F9 cell extracts than in HeLa cell extracts. Mixing experiments showed that the EC cell phenotype is dominant, supporting the hypothesis that an E1A-like positive activator of transcription is present in the EC cells. Recent work (La Thangue et al., 1990) has identified activators of E2 transcription which are down regulated on differentiation of F9 EC cells. These factors are known as DRTF1 and 2 (for differentiation regulated transcription factors); it is currently unclear whether DRTF1 is the murine equivalent of the HeLa cell factor E2F. These factors may mediate the effects of the E1A-like activity to allow EC cell activation of E2 transcription.

The E1A-like activity does not seem to be an artefact of the EC cell system as E1A-independent transactivation of the E2 promoter also occurs in uninfected mouse oocytes and preimplantation embryos,
but not in postimplantation embryos (Suemori et al., 1988; Dooley et al., 1989). As the papovaviruses are subject to transcriptional repression by the adenovirus early gene E1A product (Borrelli et al., 1984), and are transcriptionally inactive in EC cells, it has been suggested that the EC cell E1A-like factor may also be able to fulfill the negative role of E1A as a repressor of transcription. This hypothesis is supported by the observation that the F9.1 polyomavirus mutant which is not subject to F9 cell repression is also not subject to E1A repression (Hen et al., 1986). The E1A-like activity has been postulated to play a central role in the maintenance of the undifferentiated phenotype of embryonic stem cells. In support of this Young et al. (1989) have shown that E1A can down-regulate the expression of the type IV collagen and plasminogen activator genes in the differentiated derivatives of F9 cells, returning transcription to undifferentiated levels. They therefore propose that the EC cell E1A-like activity is responsible for the low levels of transcription in undifferentiated cells. Other cellular genes such as the α-fetoprotein and endo A genes which are relatively inactive prior to differentiation (Cremisi and Duprey, 1987; Vogt et al., 1988) may also be under the control of the E1A-like activity. An improved understanding of the mechanism of action of both E1A and the E1A-like activity will doubtless throw further light on the control of transcription in early embryonic cells.

1.5 RETROVIRUSES

The retroviruses are a group of RNA viruses which have attracted considerable interest in many different branches of biology. They are characterized by their ability to reverse transcribe RNA, thus forming a DNA copy of the genome, the provirus, which integrates into host chromosomal DNA. They were first identified by their pathogenic effects, and were later shown to be the primary transforming agents responsible for causing a variety of types of tumour. Recently human immunodeficiency virus, the causative agent of AIDS, was found to be a retrovirus. Retroviruses have proved useful as model systems for the study of fundamental biological problems, such as the regulation
of gene expression, and as vectors for genetic manipulations.

1.5.1. Life cycle.

Each retroviral particle consists of two strands of viral RNA wrapped in a core of viral protein which is surrounded by an envelope derived from the membrane of the previous host cell and studded with viral glycoproteins. Infection is initiated when the viral envelope binds to a target cell membrane. A murine cDNA has been isolated (Albritton et al., 1989) which encodes a putative receptor for the Moloney murine leukaemia virus (Mo-MuLV). The gene product is a multiple membrane-spanning protein, capable of conferring susceptibility to virus infection. Once bound to the cell membrane the virus is believed to enter the cell via receptor mediated endocytosis, and once within the cell the enveloped particle is converted to an enzymatically active nucleoprotein complex ready for reverse transcription.

All retroviruses have a similar genomic structure in which the central coding sequences are surrounded by terminal control sequences. The basic structure of Mo-MuLV, a typical C-type retrovirus, is shown in Fig. 1.2, and the functions of each region of the genome are summarized in Table 1.1. The virally encoded reverse transcriptase produces viral DNA using the two RNA copies as templates, and makes two transfers of nascent DNA between the templates during the process. Negative strand DNA synthesis is primed from a tRNAPro molecule which interacts with the primer binding site (PBS). A purine rich site has been implicated in priming positive strand DNA synthesis but the nature of the primer is unknown. The U3 and U5 sequences become duplicated during formation of the double-stranded viral DNA copy to produce a long terminal repeat (LTR) sequence at each end of the genome.

Circular forms of retroviral DNA containing 1 or 2 copies of the LTR have been observed in cells and it has been suggested that integration proceeds via circular intermediates. However, it has also been shown that linear Mo-MuLV DNA can be integrated directly into the host genome (Fujiwara and Mizuuch, 1988). The observation of an
integration intermediate in which 3’ ends of the LTR sequence are
joined to target DNA while 5’ ends remain free suggests that
integration proceeds via a similar mechanism to that used by
retrotransposons (reviewed by Grandgenett and Mumm, 1990). A
nucleoprotein complex capable of integration in vitro has been
isolated (Bowerman et al., 1989). It consists only of viral DNA,
capsid protein and presumably the integration protein (IN) encoded by
the 3’ terminus of the viral pol gene, showing that no host proteins
are required for integration. This result was confirmed by Katz et
al. (1990), who showed that the avian sarcoma-leukosis virus IN
protein is both necessary and sufficient for integration in vitro.

Analysis of proviral junctions has shown that the provirus is
usually within 500 bp of a DNaseI-hypersensitive site (Vijaya et al.,
1986; Rohdewohld et al., 1987). As DNaseI-hypersensitive sites are
characteristic of transcribing DNA it is likely that proviral
integration favours transcriptionally active chromatin. The mechanism
of proviral integration can produce insertion mutations which
occasionally activate proto-oncogenes leading to cellular
transformation. Retroviruses can also capture or "transduce" proto-
oncogenes to become transforming agents. This may involve mutation of
the proto-oncogene or its inappropriate expression (reviewed by Weiss
et al., 1982).

The retrovirus replication system allows any sequences placed
between the terminal control sequences to be inserted into the genome
in the form of a provirus, providing a very useful vector system
(reviewed by Gilboa et al., 1986). Retroviral vectors can be used as
lineage markers to label the genome and allow the identification of
daughter cells (reviewed by Price, 1987), or to cause mutations, to
overexpress genes and potentially to correct genetic defects
(reviewed by Friedmann, 1989).

The life-cycle is completed by transcription of viral RNA under the
control of the promoter and enhancer sequences in U3. The RNA is a
template for the translation of viral gene products, and is also the
viral genome which combines with the viral proteins under control of
the packaging signal psi to form new retroviral particles. The
particles bud off from the host cell and are ready to initiate new
TABLE 1.1 The functions of each region of the Mo-MuLV genome

<table>
<thead>
<tr>
<th>Region</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>A 60 nucleotide sequence repeated at each end of the genome. Involved in transfer of the nascent DNA strand during reverse transcription.</td>
</tr>
<tr>
<td>U5</td>
<td>A unique sequence of 60 nucleotides separating R and the primer binding site.</td>
</tr>
<tr>
<td>PBS(-)</td>
<td>The &quot;primer binding site&quot;. The binding site for tRNA&lt;sub&gt;Pro&lt;/sub&gt; which primes negative strand DNA synthesis.</td>
</tr>
<tr>
<td>psi</td>
<td>The packaging signal, required for packaging of genomic RNA into viral particles, and involved in dimerization of the two RNA molecules found in each particle.</td>
</tr>
<tr>
<td>gag</td>
<td>One of the two major RNA species, a full length 35S RNA, is translated to produce the Pr65 gag gene product, the precursor to the internal structural proteins of the viral particle.</td>
</tr>
<tr>
<td>pol</td>
<td>The 35S RNA also produces the Pr180 gag-pol gene product. This is cleaved to give the pol gene product reverse transcriptase, the integration protein (IN), and the protease required for maturation of structural proteins (Yoshinaka et al., 1985).</td>
</tr>
<tr>
<td>env</td>
<td>The other transcription product, a 22S RNA derived by alternative splicing, is translated to give Pr70 the env gene product, the surface glycoprotein on the viral envelope.</td>
</tr>
<tr>
<td>PBS(+)</td>
<td>A purine rich region 5' of the initiation site of positive strand DNA synthesis, implicated as a primer binding site.</td>
</tr>
</tbody>
</table>
U3 A unique sequence separating R and PBS(+), which has two copies in the provirus. U3 contains the promoter and enhancer of viral transcription and includes two 75 bp direct repeat sequences.

A_{200} A poly-A sequence of approximately 200 residues occurs at the 3' end of the genome. Addition of the poly-A tail is post-transcriptional.
Fig. 1.2. Structure of the Moloney murine leukaemia virus genome. For definitions of each region see Table 1.1. (Taken from Weiss et al., 1982; 1985.)
Structure of the Moloney Murine Leukaemia Virus Genome

Viral RNA

```
<table>
<thead>
<tr>
<th>R</th>
<th>U₅</th>
<th>PBS(−) ψi</th>
<th>gag</th>
<th>pol</th>
<th>env</th>
<th>PBS(+)</th>
<th>U₃</th>
<th>R</th>
</tr>
</thead>
</table>
```

Protein Products

- Pr65
- Pr180
- Pr70

Proviral DNA

```
<table>
<thead>
<tr>
<th>U₃</th>
<th>R</th>
<th>U₅</th>
<th>gag</th>
<th>pol</th>
<th>env</th>
<th>U₃</th>
<th>R</th>
<th>U₅</th>
</tr>
</thead>
</table>
```

Cellular DNA

```
72 bp Direct repeats
```

Start
Fig. 1.3. Retroviral life cycle.
Retroviral Life Cycle

Adsorption → Uncoating → Reverse Transcription → DNA → DNA Uncoating → Integration → Provirus → Transcription → mRNA → Nucleus → Assembly → Translation → mRNA → Translation → Viral Proteins → Budding → Cell
infections (Fig. 1.3).

1.5.2. Moloney murine leukaemia virus.

Mo-MuLV is a C-type retrovirus which was obtained by the passage of a cell-free extract of the Sarcoma 37 transplantable tumour in neonatal mice (Moloney, 1960). Its general pathology includes thymic leukaemia, disseminated lymphosarcoma or lymphatic leukaemia and hepatosplenomegaly. The Moloney murine sarcoma virus (Mo-MSV) is a replication defective derivative of Mo-MuLV in which part of the $env$ open reading frame has been replaced with the $mos$ oncogene. Mo-MSV was isolated from a sarcoma caused by injecting a BALB/c mouse with Mo-MuLV (Moloney, 1966).

1.5.3. Transcriptional control of Mo-MuLV.

The sequences mediating transcription of Mo-MuLV and Mo-MSV, and the factors acting at these sites have been investigated in a variety of cell types. The U3 sequences of the two viruses, which direct transcription, are almost identical; Mo-MSV has two single bp additions and three single bp deletions with respect to Mo-MuLV (Hilberg et al., 1987). Results relating to the control of transcription obtained using one virus can generally be assumed to apply to the other. Laimins et al. (1982) showed that the 72 bp direct repeat sequences of the Mo-MSV LTR act as an enhancer of transcription. When the SV40 enhancer was replaced by the Mo-MSV 72 bp repeats the mutant SV40 virus was still competent for infection of murine cells, and T-antigen production remained at wild-type levels. On infection of monkey kidney cells T-antigen production fell by 5-fold with respect to wild-type SV40, reflecting species-specificity of the Mo-MSV enhancer.

This phenomenon was further investigated by transient transfection experiments, using constructs in which the chloramphenicol acetyltransferase (cat) gene coding sequences are under the control of the SV40 early gene promoter and the SV40 or Mo-MSV enhancer. The results in murine and monkey cell lines reflected those of the infection
experiments - in monkey CV-1 cells the SV40 enhancer was 5-fold more efficient than the Mo-MSV enhancer, and in mouse L cells the Mo-MSV enhancer was 3-fold more efficient than the SV40 enhancer. Laimins et al. (1982) noted a region of shared homology between the viral LTRs; Mo-MSV has sequences corresponding to the SV40 enhancer core motif, suggesting that the two different enhancers may act at least in part via a common mechanism. Levinson et al. (1982) also replaced the SV40 direct repeats with the Mo-MSV direct repeats and produced a viable mutant virus. Spandidos and Wilkie (1983) confirmed the species-specific activity of Mo-MSV by demonstrating that it has a higher activity in murine cells than in rat, bovine or human cell lines.

Augereau and Wasylyk (1984) used recombinants of the Mo-MuLV or SV40 enhancer sequences linked to the conalbumin promoter or the SV40 early promoter in transient transfection assays, and showed that these sequences act in a distance and orientation independent fashion, consistent with their definition as enhancers. However, the relative efficiencies of the SV40 and Mo-MuLV enhancers in these experiments were not consistent with the results of Laimins et al. (1982); the Mo-MuLV enhancer was found to be 25-fold less efficient than the SV40 enhancer in mouse L cells as opposed to 3-fold more efficient. The experiment was repeated using Mo-MSV rather than Mo-MuLV sequences with the same result. There is no obvious explanation for the inconsistency but it does exemplify the difficulties inherent in this type of experiment. These include variations in cell culture maintenance, DNA preparation and precise details of plasmid construction which can all have profound effects on transfection assay results.

In more recent work Laimins et al. (1984) further characterized the Mo-MSV enhancer. The direct repeat sequences were again shown to act in an orientation and distance independent fashion consistent with their definition as enhancers. The promoter distal repeat was about 3-fold more active than the promoter proximal repeat, the difference in activity is presumably mediated by the 4 bp changes between the two repeats. It was also found that the GC-rich sequence between the direct repeats and the start of transcription could be effectively replaced by the GC-rich 21 bp repeats of the SV40 early gene.
Graves et al. (1985a) further delineated the transcriptional control signals of Mo-MSV, using constructs in which the thymidine kinase (tk) gene coding sequences of Herpes simplex virus (HSV) were placed under the control of various U3 deletions. Steady state transcription rates were analyzed in micro-injected *Xenopus laevis* oocytes and transfected mouse fibroblasts, and transient transcription rates were analyzed in mouse fibroblasts. Three separate control sequences were identified by this approach:

(i) A TATA homology at -21 to -31 bp (numbered with respect to the start of transcription). This site is essential for accurate initiation (*i.e.* the production of transcripts with the correct 5' terminus) in *Xenopus* oocytes.

(ii) The distal signal, -31 to -84 bp. This sequence is essential for transcription in *Xenopus* oocytes; in mouse fibroblasts it is not essential but increases the rate of transcription. The sequence includes a CCAAT box homology, which was later shown to interact with the rat liver CCAAT binding protein (CBP) (Graves et al., 1986).

(iii) The enhancer, -156 to -365 bp, which includes the direct repeat sequences. The entire sequence increases transcription 7-fold in transient transfection assays and 4-fold in stably transfected mouse fibroblasts. However this sequence has no effect on transcription in *Xenopus* oocytes, again illustrating the species-specificity of the enhancer. An accompanying paper (Graves et al., 1985b) showed that the infection of mouse L cells by HSV facilitates activity of the distal signal. These results led Graves et al. (1986) to the hypothesis that Mo-MSV has two alternative modes of transcriptional activation, one via the direct repeat sequences and another via the distal signal.

The Mo-MSV enhancer sequences were shown by DeFranco et al. (1985) to facilitate activation by the synthetic glucocorticoid hormone dexamethasone in transient transfection experiments. The glucocorticoid receptor was shown by footprinting studies to interact with three separate sites in U3, one in each direct repeat and one further upstream. Each of the sites has homology with the consensus glucocorticoid receptor binding site (GRE). Miksicek et al. (1986)
obtained similar results in transfection and footprinting experiments. They then analyzed the ability of each of the three potential GREs to mediate dexamethasone activation of transcription in transient transfections. Sites within the direct repeat sequences were found to be functional but the third site was non-functional. However, it is possible that this site cooperates with the others in vivo. One interesting result of these experiments was that only certain restriction fragments containing the GRE were functional, suggesting a requirement for interactions with other enhancer elements. Sap et al. (1989) reported that the Mo-MuLV U3 region also contains a thyroid hormone response element (TRE), facilitating activation of transcription by c-erb-A protein.

Speck and Baltimore (1987) attempted to further define the transcription factor binding sites in the 72 bp direct repeats of Mo-MuLV by methylation interference footprinting and the gel retardation technique. Multiple proteins in WEHI 231 (a murine B cell line) nuclear extracts were able to interact with the Mo-MuLV sequences. In order to distinguish between these proteins it was necessary to partially fractionate the extract by column chromatography. Six separate binding activities were revealed. Three of these bound to sites previously characterized:

(i) The GRE site occurring once in each 72 bp repeat.
(ii) The enhancer core motif with homology to that of SV40 (Weiher et al., 1983) also occurring once in each repeat.
(iii) Sites with homology to the consensus NF-1 binding site (Nagata et al., 1982) occurring twice in each repeat.

The other three binding sites had not previously been recognized and were designated LVa, b and c.

In vitro techniques have also revealed a binding site for the proto-oncoprotein ets-1 (Gunther et al., 1990). The site lies within the distal signal around -45 bp upstream of the transcriptional start. Mutation analysis has shown that disruption of the binding site reduces expression from the LTR by 15 to 20-fold in cultured mouse thymocytes.

There are a number of murine leukaemia viruses related to Mo-MuLV, but with different disease forming specificities. A large body of
evidence has accumulated that suggests the tissue-specific transcription afforded by different U3 sequences is responsible for these varying disease specificities. Li et al. (1987) showed that the U3 region of Friend murine leukaemia virus (Fr-MLV) confers the ability to produce erythroleukaemia upon Mo-MuLV. Bosze et al. (1986) and Short et al. (1987) used transient transfection assays to show that the transcriptional activity of each virus correlates with its pathogenicity in a given cell line. Thiesen et al. (1988) compared the enhancer sequences within the U3 regions of Mo-MSV and Fr-MLV and noticed a repeated sequence of 5'TGCC 3' occurring four times in Mo-MSV but mutated at each of the four equivalent sites in Fr-MLV. Sequential mutation of the sites showed that they confer T-cell specific transcription. Manley et al. (1989) used gel retardation and footprinting experiments to investigate the proteins binding to Fr-MLV and Mo-MuLV LTRs in various cell lines, and discovered both shared and unique binding sites. Wolff et al. (1990) noted that the capacity of Mo-MuLV to induce myeloid tumours requires both a structural element, which is absent in Fr-MLV, and a control element located upstream of the direct repeat sequences in U3. The latter element was shown to be required for myeloid but not lymphoid tumour production by insertion mutagenesis. Numerous other examples of the correlation between specificity of transcription and pathology have been reported for other retroviruses (e.g. Celander and Haseltine, 1984; Vogt et al., 1985; Short et al., 1987; Friel et al., 1990).

The determinants of viral pathogenesis have been further localized to single protein-binding sites in the Mo-MuLV enhancer (Speck et al., 1990b) - the latent period of disease onset is increased by mutations in the LVb, core, NF-1 or two or more GRE sites. This result upon mutation of the GRE elements suggests that they are functionally significant, although Overhauser and Fan (1985) did not observe activation of transcription from Mo-MuLV by the synthetic glucocorticoid dexamethasone. This was an unexpected result as the Mo-MuLV GREs are identical in sequence to the Mo-MSV GREs which do mediate activation by dexamethasone. Mutations in the core and LVb sites can also change the disease-specificity of Mo-MuLV, allowing
the induction of erythroleukaemias rather than lymphomas. This type of enhancer-specific effect may be mediated in one of two ways:-
a) by high transcriptional activity producing high levels of a viral gene product involved in tumour formation, or alternatively
b) a more active enhancer may have a greater probability of activating neighbouring proto-oncogenes.

However, it should be noted that Paludan et al. (1989) reported that following integration the proviral flanking sequences frequently override the transcriptional control sequences of the virus.

1.5.4. The non-permissivity of stem cells for Mo-MuLV.

Jaenisch et al. (1975) found that Mo-MuLV generally produces silent infections of preimplantation embryos, although integration of the provirus into the host genome does occur. Later it was observed by Teich et al. (1977) that three embryonal carcinoma (EC) stem cell lines derived from strain 129 mice are not susceptible to infection by Mo-MuLV. RNA viruses from different families and the DNA virus vaccinia are able to productively infect these cells. One of the three cell lines, PSA 4, can be differentiated in vitro to form cystic embryoid bodies, which are permissive for retrovirus infection. Vesicular stomatitis virus (VSV) is able to infect EC cells; when its capsid is replaced with that of Mo-MuLV it remains infectious, showing that the Mo-MuLV virus particle is capable of entering EC cells. Nucleic acid hybridization experiments indicate the presence of proviral DNA in the EC cells, but the absence of viral RNA suggests that the block to infection lies at the level of transcription. Differentiation of previously infected cells did not result in virus production, showing that the developmental stage at which infection occurs determines the outcome of the infection. Peries et al. (1977) also observed that murine ecotropic C-type retroviruses do not multiply in strain 129 mouse EC cells but that they do multiply in differentiated derivatives of those cells.

Recent work (Savatier et al., 1990) has more accurately pinpointed the point at which embryonic cells become permissive for retroviral infection. Blastocysts and postimplantation embryos were infected
with recombinant viruses and LTR mediated expression was only observed in mesoderm and definitive ectoderm. Permissiveness is apparently obtained immediately upon differentiation of epiblast during gastrulation of the mouse embryo.

Stewart et al. (1982) and Niwa et al. (1983) investigated the methylation state of proviral DNA in EC cells. Using F9 and EC-A1 (a PCC4 derivative) EC cells respectively, they found that the proviral DNA was highly methylated and that demethylation allowed viral replication to proceed. In cells that had been infected when in a differentiated state the methylation level of the provirus was normal. Jahner et al. (1982) obtained similar results on comparison of retroviral infection of preimplantation and postimplantation mouse embryos. When the embryos were infected preimplantation 70% to 90% of CpG residues in the provirus became methylated. Conversely, when infection occurred postimplantation 80% of these residues remained unmethylated. Although these results led to the hypothesis that transcription of proviruses is blocked in embryonic stem cells by hyper-methylation, studies of methylation kinetics (Gautsch and Wilson, 1983) showed that the provirus does not become hyper-methylated until at least 8 days post-infection. It seems that methylation is a result rather than a cause of transcriptional inactivation. Attention was therefore turned to the control of proviral transcription initiation in early embryonic cells.

Linney and co-workers showed that the polyoma virus is unable to replicate in EC cells due to inactivity of the viral enhancer (Fujimura et al., 1981). They postulated that the block to Mo-MuLV provirus transcription in embryonic cells could be mediated in an analogous fashion. In order to test this prediction they performed transient transfection experiments using constructs in which the cat gene coding sequences were under the control of wild-type and mutant forms of the Mo-MuLV U3 sequences (Linney et al., 1984). The wild-type U3 sequences facilitated CAT expression in a fibroblast cell line, but not in F9 EC cells. Deletion of the direct repeat sequences in the viral LTR rendered transcription virtually undetectable by a CAT assay in both cell types. The deleted sequences were then replaced with the enhancer of a mutant polyoma virus, Py F101,
capable of replication in EC cells (Fujimura et al., 1981). The Py F101 enhancer increased transcription rates in the fibroblast cells by 2 to 3-fold and also allowed transcription to occur at an equivalent rate in the EC cells. Similar results were obtained if mutant viruses were used to investigate infectivity; if the Mo-MuLV enhancer is replaced with that of Py F101 a productive infection of F9 EC cells can occur. These results suggest that at least part of the block to Mo-MuLV infection of F9 cells is caused by inability of the enhancer sequence to function in EC cells.

The results of Linney et al. (1984), discussed above, suggest that the Mo-MuLV enhancer and promoter are both virtually inactive in EC cells. However, more detailed studies using Mo-MSV (Gorman et al., 1985a; 1985b; Sassone-Corsi et al., 1985) have shown that the promoter alone does have weak activity in EC cells and that this activity is repressed by the enhancer. Gorman et al. (1985a) showed that a construct in which the cat gene is placed under the control of the Mo-MSV enhancer and promoter has a very low expression level in F9 EC cells, only about one-thirteenth of the expression level from an equivalent SV40 early gene construct. If the enhancer sequences are deleted the expression level increases by about ten-fold, suggesting that the Mo-MSV enhancer contains an element able to mediate down-regulation of transcription. If the promoter sequences are also deleted CAT activity becomes undetectable. In the experiments performed by Linney et al. (1984) Mo-MuLV and an enhancer deletion construct derived from it are apparently inactive in EC cells. The different results obtained using Mo-MuLV and Mo-MSV may represent a genuine difference in the transcriptional control mechanisms of the two viruses, but in view of the high level of sequence similarity between the viral LTRs it is perhaps more likely that the more complex results obtained by Gorman et al. (1985a; 1985b) reflect a higher transfection efficiency, allowing small variations in CAT activity to be more easily detected. Alternatively, the discrepancy may reflect differences between the cell lines used in each experiment; it is possible that differing culture conditions of the F9 cells may have resulted in changes of phenotype.

Further evidence for the existence of a negative regulatory factor
in EC cells comes from co-transfection experiments (Gorman et al., 1985a, 1985b). If F9 EC cells are co-transfected with Mo-MSVcat and a construct in which the neomycin phosphotransferase (neo) gene is placed under the control of the Rous sarcoma virus (RSV) LTR, the level of CAT expression increases in proportion to the concentration of RSVneo. Co-transfection has no effect on the expression of constructs in which the Mo-MSV enhancer has been deleted. These results suggest that a negative regulatory factor is being titrated out of the EC cells allowing an increase in transcription initiation, and also imply that the factor must recognize the RSV LTR. As an RSVcat construct is expressed with high efficiency in EC cells the repressor must interact with RSV in a non-functional manner. Gorman et al. (1985b) also showed that the SV40 enhancer is inactive in EC cells. The SV40 promoter alone directs a low level of transcription in EC cells and this level is unaffected by addition of the enhancer sequence. If the SV40 enhancer is replaced by the Mo-MSV enhancer the low level of CAT expression directed by the SV40 promoter in EC cells is completely abolished.

Loh et al. (1987) performed similar experiments using Mo-MuLV in F9 EC cells and a second EC cell line PC13. Constructs were used in which the neo coding sequences are under the control of Mo-MuLV enhancer and promoter, or the promoter alone. Both the wild-type and deleted constructs directed an equivalent low level of neo transcription in EC cells, whether monitored in transient or stable transfectants. The discrepancy between these results and those of Gorman et al. (1985a, 1985b) and Sassone-Corsi et al. (1985) can once again be explained by differences between Mo-MuLV and Mo-MSV, differences in cell lines or differences in transfection efficiency.

Loh et al. (1987; 1988) identified a negative regulatory element downstream of the transcriptional start site. Transfection studies in EC cells using a cloned provirus showed further down-regulation of expression in comparison to the Mo-MuLV U3neo construct described above (Loh et al., 1987). A negative regulatory activity was further localized by deletion studies to a 29 bp intergenic sequence which includes the tRNA primer binding site (PBS). Further studies (Loh et al., 1988) showed that the PBS inhibits expression in tranfection
assays of EC cells but not C2 myoblast cells.

The tRNA PBS was also identified by Barklis et al. (1986) as a negative regulator of provirus expression. This group selected for the few proviruses which give a productive infection of F9 EC cells; generally this was found to be a function of the flanking genomic sequences, but in one provirus a single point mutation had occurred within the PBS which conferred the ability to express in F9 EC cells. The mutant virus was named B2 and shown to produce proviral message in F9 cells, although still at a 20 to 100-fold lower efficiency than in fibroblast cell lines. Taketo et al. (1989) used a similar approach to isolate mutant viruses capable of expression in EC cells. They too isolated viruses with mutations in the PBS, but also found mutations in the splicing donor site could confer EC cell expression, suggesting a splicing deficiency in EC cells.

Recently, mutagenesis studies have further defined the negative regulatory element (NRE) to an 18 bp region spanning the tRNA PBS (Loh et al., 1990). The NRE is not only active in an inverted orientation but also, to some extent, when moved upstream of the transcriptional start. The region has been shown to interact with an EC cell-specific protein which is a potential repressor of transcription.

It seems from these results that one mechanism by which provirus expression is blocked in EC cells is indeed mediated by the PBS sequence. However, this does not appear to be the sole mechanism of repression. The transfection experiments of Gorman et al. (1985a; 1985b) and Sassone-Corsi et al. (1985) suggest a negative role for the viral enhancer. This hypothesis is supported by the isolation of further mutant viruses capable of expression in EC cells.

The myeloproliferative sarcoma virus (MPSV) was isolated by Chirigos et al. (1968) on serial passage of a tumour induced in newborn mice by uncloned Mo-MSV with Mo-MuLV as a helper virus. When cloned and sequenced (Kollek et al., 1984; Stacey et al., 1984) MPSV was shown to be similar in construction to Mo-MSV, consisting only of sequences related to Mo-MuLV or to the mouse mos proto-oncogene which has replaced the pol and env genes of Mo-MuLV. The deletion renders the virus incompetent for replication and therefore it requires a
helper virus such as Mo-MuLV. Sequencing revealed several point mutations in the U3 region of the virus. On comparison with Mo-MSV there were seven single bp deletions, two single bp mutations and two single bp additions in MPSV. Ostertag's group observed that MPSV has a wider host range than Mo-MuLV - it not only transforms fibroblasts but also causes a myeloproliferative syndrome involving both the myeloid and erythroid compartments of the haematopoietic system (Klein et al., 1981; Stacey et al., 1984). In vitro stimulation of the proliferation of erythroid early precursor cells (BFU-E), myeloid precursor cells (CFU-C) and stem cells (CFU-S) by MPSV has also been observed. Stocking et al. (1985) used mutant viruses combining sequences from MPSV and Mo-MSV to define those sequences responsible for this increased host range. Both the U3 sequence and the oncogene were shown to play a vital role in determining pathogenicity.

Ostertag and co-workers (Franz et al., 1986; Seliger et al., 1986) used a derivative of MPSV, encoding the neo resistance gene, to demonstrate that its host range also includes embryo-derived stem cells (ES cells) and F9 EC cells, but not PCC4 EC cells. Transfer of G418-resistance, a function of the neo gene product, by MPSV was almost as efficient to F9 cells as to fibroblasts, but was about one thousand-fold less efficient to PCC4 cells. The rare PCC4 cells which did express neo were induced to differentiate by retinoic acid treatment and then superinfected with Mo-MuLV to allow rescue of the virus. Two viruses were rescued and termed PCMV-5 and -6 for PCC4 cell-passaged NeoR-MPSV; each virus transferred G418 resistance to both F9 and PCC4 cells. The PCMV virus was later cloned and sequenced (Hilberg et al., 1987). Two deletions were found in PCMV, one of 75 bp encompassing a single direct repeat of the LTR, and one of about 200 bp at the border between the viral and neo sequences. In addition to the deletion there are eleven single bp changes in the LTR between MPSV and PCMV, most of which are reversions to the Mo-MSV or Mo-MuLV sequence; only one point mutation unique to PCMV lies in the U3 region, at -345 bp.

By replacing the U3 region of MPSV with that of PCMV, Hilberg et al. (1987) showed that the U3 region of PCMV is sufficient to allow infection of PCC4 EC cells. The PCMV enhancer region was also active
in PCC4 cells in transient transfection experiments when used in place of the Mo-MuLV enhancer in a Mo-MuLV-cat construct. The increased host range of PCMV is therefore a function of increased initiation of transcription facilitated by changes in the enhancer sequences. Weiher et al. (1987) compared the effects of the MPSV mutation with those of the B2 mutation (Barklis et al., 1986) in both F9 and PCC4 EC cells. Gene expression of the MPSV mutant is about 100-fold higher in F9 cells than in PCC4 cells. However, the B2 mutant expresses at equivalent levels in the two cell lines, the absolute level being intermediate between that of MPSV in F9 cells and in PCC4 cells. If the MPSV upstream sequences are linked to the B2 downstream sequences, expression in both cell types is increased to the level attained by MPSV in F9 cells. This result suggests that the inactivity of MPSV in PCC4 cells is not mediated by the enhancer sequences but by the PBS sequences. The MPSV PBS has in turn been reported to be less inhibitory to expression in F9 EC cells than the Mo-MuLV PBS (Feuer et al., 1989). However, any inhibitory effect of the PBS sequence in EC cells can clearly be counteracted by increased activity of the enhancer. This is exemplified by a construct in which the PCMV enhancer confers the ability to be transcribed in PCC4 cells upon Mo-MuLV (Hilberg et al., 1987).

Reisman and Rotter (1989) investigated whether another type of stem cells, human myeloid cells, would allow increased initiation of transcription from the Mo-MuLV LTR following differentiation. They performed transient transfection experiments using a construct in which the cat gene coding sequences are under control of the SV40 promoter and the Mo-MuLV enhancer in a pluripotential human myelocyte cell line, HL-60. They found that the Mo-MuLV enhancer functions poorly in the undifferentiated cells but is induced by about 10-fold upon terminal differentiation of the cells.

Attempts have been made to determine which transcription factors are responsible for the inactivity of the Mo-MSV LTR in stem cells. Speck and Baltimore (1987) identified, by methylation interference and gel retardation experiments, six different activities from WEHI 231 cells which interact with the Mo-MuLV enhancer. They then compared the abundance of these activities in WEHI 231 cells, NIH 3T3
cells (murine fibroblasts) and in three stem cell lines, F9, PCC4, and MEL (murine erythroid leukaemia) cells. Gel retardation experiments suggest that the factors which binds to the enhancer core motif and NF-1 site are at lower concentration in the stem cells than in WEHI 231 or NIH 3T3 cells. This result suggests that low transcription factor concentrations in stem cells do contribute to the inactivity of the Mo-MuLV enhancer. However, this cannot explain the negative effect on transcription exerted by the enhancer (discussed above). Flamant et al. (1987) approached the problem by comparing the factors which interact with the various restriction fragments of Mo-MuLV U3 from both PCC4 EC cells and their differentiated derivatives. They found that EC cells contain 10-fold higher concentration of a factor, EPBF, which interacts with the distal signal between -59 and -87 bp (numbered with respect to the start of transcription) in the vicinity of the CAAT box homology. EPBF is postulated to be a repressor of transcription but no functional evidence has yet been presented to support this hypothesis.

In summary, it appears that the block to Mo-MuLV infection of stem cells is mediated via a block to transcription. However the precise mechanism of the block remains unclear. Transcription may not be initiated due to lack of positive acting transcription factors, due to the presence of repressors of transcription, or due to a combination of both. The nature of the factors involved may vary from one type of stem cell to another.

1.6 AIMS OF THE PROJECT

Past experience has shown that viral genes are a useful model system for studies of transcriptional control. One important, yet relatively poorly understood transcriptional control system is that of pluripotential embryonic stem cells. The papovaviruses and adenoviruses, which demonstrate regulated expression during stem cell differentiation, have been exploited in order to investigate variations in the transcription apparatus between stem cells and their more differentiated derivatives. However, less progress has
been made in understanding a stem cell-specific block to the expression of retroviruses. The aims of this project are to shed further light on the mechanisms blocking transcription of retroviruses in embryonic stem cells, and in the process to further our understanding of general transcriptional control in stem cells.

The approach taken to address these aims made use of the Moloney murine sarcoma virus (Mo-MSV) and F9 EC cells. Mo-MSV was already cloned, as were mutant forms which relieve the block to expression in a variety of EC cell lines. The F9 EC cell system was used as a model system representing early embryonic stem cells because these cells yield sufficient quantities of protein to perform in vitro experiments, and are amenable to transfection allowing in vivo studies. The approach used was to identify cis-acting sequence elements and their trans-acting DNA-binding proteins with a potential role in the control of Mo-MSV transcription in EC cells, using the gel retardation assay. Candidate sequence elements of interest were identified by comparing the sequence of wild-type Mo-MSV with that of mutant forms capable of expression in EC cells. The functional significance of the binding sites was then investigated using transient transfection assays.
CHAPTER 2

MATERIALS AND METHODS
2.1 GENERAL METHODS

2.1.1. Safety.

A laboratory coat was worn at all times, and latex gloves and eye-protection where appropriate. Toxic or hazardous chemicals were handled in a fume cupboard. $^{32}$P compounds were handled behind a 1cm perspex screen, whilst wearing two pairs of latex gloves. All $^{32}$P work was carried out in a room designated for that purpose. A radioactive monitor was used constantly, and personal exposure monitored using finger tapes. Biohazardous waste was disposed of by autoclaving.

2.1.2. Sterilization.

Plasticware such as microcentrifuge tubes and pipetteman tips was sterilized by autoclaving at 121°C, 15 pounds/inch$^2$ for 30 minutes (min). Solutions were sterilized in the same way where possible, or alternatively filter sterilized through a 0.45μm Nalgene disposable filter. Deionized distilled water was used for the preparation of all solutions. For RNA work, solutions were treated with 0.1% (v/v) DEPC (Sigma) prior to autoclaving, and glassware was baked for 24 hours (h).

2.1.3. Extractions and precipitations.

Phenol/chloroform extractions were performed in order to remove contaminating proteins from DNA solutions. Phenol (BRL) was equilibrated with 10 mM Tris (BDH) adjusted to pH 7.5 with 1 M HCl (FSA Laboratory Supplies) and stored at -70°C (long term) or at 4°C in the dark. All chloroform (May and Baker) used was a 24:1 mixture with the anti-foaming reagent isoamyl alcohol (Sigma). The solution was extracted with an equal volume of phenol : chloroform : isoamyl alcohol in the ratio 25:24:1, vortexed and microcentrifuged at 14,000 rpm (using an Heraeus sepathec Biofuge A). The upper aqueous phase was then re-extracted with chloroform, vortexed and microcentrifuged.
once again. 1/10 volumes of sodium acetate (Sigma; pH 5.2) and 2 volumes 96% (v/v) ethanol (EtOH; Hayman Ltd.) were added to the aqueous phase, and precipitation carried out on solid CO₂ for 15 min. The DNA was pelleted by microcentrifugation for 15 min, washed with 1ml 70% (v/v) EtOH, air-dried and resuspended in dH₂O.

2.1.4. Preparation of dialysis tubing.

Dialysis tubing (Medicell International) was prepared for use by boiling in 10 mM EDTA (pH 8.0; Boehringer Mannheim) for 2 h followed by autoclaving at 121°C for 30 min then stored in 50% (v/v) EtOH at 4°C.

2.1.5. Assays of nucleic acid and protein concentration.

A Uvikon 860 spectrophotometer was used to determine the concentration of nucleic acids according to the following relationships (Sambrook et al., 1989):

\[
\begin{align*}
1 \ A_{260} &= 50 \ \mu g/ml \ \text{double-stranded DNA.} \\
1 \ A_{260} &= 40 \ \mu g/ml \ \text{RNA.} \\
1 \ A_{260} &= 20 \ \mu g/ml \ \text{oligonucleotides.}
\end{align*}
\]

A pure preparation of DNA was assumed to have an A_{260} / A_{280} ratio = 1.8. A pure preparation of RNA was assumed to have an A_{260} / A_{280} ratio = 2.0. Silica cuvettes were used for absorbance readings.

Protein concentrations were measured using the Bio-Rad protein concentration assay, based upon the method of Bradford (1976). Samples were added to a 1/5 dilution of the assay, the A_{595} measured in disposable cuvettes and compared to standards of BSA (Bio-Rad) at various known concentrations.

2.2 SYNTHETIC OLIGONUCLEOTIDES

Synthetic oligonucleotides were prepared by Mike Furness and Diane Chase using an Applied Biosystems automatic synthesizer. The following oligonucleotides were used.
CAT-1: Coding sequence of chloramphenicol acetyltransferase gene from nucleotide +15 to +34 (Gorman et al., 1982b).

5’ CAATCGGTATATCGTG 3’

MSV: MSV enhancer sequence -363 to -333, with respect to the transcriptional start (Dhar et al., 1980).

5’ AGAGTCACTAGCTAAGAGAAGAAGA 3’
3’ TCTAGCTACAGTGATCTTCGTCTCTTC 5’

PCMV: PCMV enhancer sequence at equivalent position to MSV enhancer sequence -363 to -333 (Hilberg et al., 1987).

5’ AGAGTCACTAGCTAAGAGAAGAAGA 3’
3’ TCTAGCTACAGTGATCTTCGTCTCTTC 5’

Sp1 WT: PCMV enhancer sequence from -176 to -157 (Hilberg et al., 1987).

5’ GATGGTCCTACCTGCTACGC 3’
3’ TCTAGCTACAGTGATCTTCGTCTCTTC 5’

Sp1 mut: MSV enhancer sequence from -176 to -157 (Dhar et al., 1980).

5’ GATGGTCCTACCTGCTACGC 3’
3’ TCTAGCTACAGTGATCTTCGTCTCTTC 5’

NF-1 WT: MSV enhancer sequence from -333 to -309 (Dhar et al., 1980).

5’ GAGAACAATGGTGGCCATGGACAGAGA 3’
3’ CTCTGTGGATCTTCTTATATCGGATGGATGCC 5’

NF-1 mut: PCMV enhancer sequence from -333 to -309 (Hilberg et al., 1987).

5’ GAGAACAATGGTGGCCATGGACAGAGA 3’
3’ CTCTGTGGATCTTCTTATATCGGATGGATGCC 5’

Py WT: Polyomavirus sequence from 5103 to 5135 (Piette and Yaniv, 1987).

5’ GCCAGTCCTAGCTATTCCTGCTTGA TTGATC 3’
3’ CTTAGGCTACCTACCTT GCTGCTGACGGCAGCTCGTACGATGCTAGCTACCC 5’

Py mut: As above with a point mutation at 511.

5’ GCCAGTCCTAGCTATTCCTGCTTGA TTGATC 3’
3’ CTTAGGCTACCTACCTT GCTGCTGACGGCAGCTCGTACGATGCTACCC 5’


5’ GATCTAACCTAGCTATTCCTGCTTGA TTGATC 3’
3’ ATTTAGGCTACCTACCTT GCTGCTGACGGCAGCTCGTACGATGCTACCC 5’

(Single base-pair (bp) variations are emboldened).
Oligonucleotides were heat deprotected in 35% (v/v) ammonia (BDH) at 55°C for 6 to 12 h, then EtOH precipitated. Double-stranded (ds) oligonucleotides were prepared by annealing equimolar concentrations of complementary single-stranded (ss) oligonucleotides in dH₂O by heating to 90°C and cooling to room temperature over approximately 3 h. The success of the annealing reaction was tested by electrophoretic analysis. 200ng of dsDNA products or ssDNA reagents were suspended in 5µl of 1 x agarose gel loading buffer (Table 2.1) and electrophoresed at 10 V/cm through a 2% (w/v) agarose (Sigma) gel in the presence of 0.5 µg/ml ethidium bromide (EtBr; Sigma) for 1 to 2 h in 1 x TBE buffer (Table 2.1). The dsDNA chelates EtBr more efficiently than ssDNA and therefore fluoresces more strongly under UV illumination at 302nm (using a UV Products Ltd. transilluminator); also the migration rate of dsDNA is generally slightly different to that of its ssDNA components. Oligonucleotides were stored at -20°C in dH₂O.

2.3 RESTRICTION ENDONUCLEASE DIGESTION

2.3.1. Diagnostic digestions.

Diagnostic digestions used 1-5µg of plasmid DNA and approximately 3 units of enzyme for every µg of DNA (at no time did the volume of enzyme exceed 10% of the reaction volume) in a final volume of 20µl of 1 x restriction buffer. The buffers used were chosen according to manufacturer’s recommendation and are listed in Table 2.2. Reactions were incubated for a minimum of 1 h at 37°C. 1µl of 5 x agarose gel loading buffer was added to 4µl of the reaction products which were then analyzed by electrophoresis through an agarose mini-gel (8cm x 10cm) containing 0.5 µg/ml (w/v) EtBr at 10 V/cm for 2 h in 1 x TBE buffer (Table 2.2). Fragments of less than 1 kb were generally analyzed on 1.4-2% (w/v) gels, larger fragments on 0.6-1.4% (w/v) gels. Reaction products were visualized using a UV transilluminator fluorescing at 302nm. Size markers of lambda DNA (BRL) digested with HindIII, or pBR322 DNA digested with HinfI were electrophoresed beside the samples. Results were recorded by photography using HP5.
TABLE 2.1 Electrophoresis buffers

5 x Agarose gel loading buffer.
- 0.25% (w/v) bromophenol blue (BDH)
- 0.25% (w/v) xylene cyanol (BDH)
- 25% (w/v) Ficoll 400 (Sigma)

SDS-PAGE loading buffer.
- 4% (w/v) SDS (BDH)
- 20% (v/v) glycerol (FSA Laboratory Supplies)
- 0.04% (w/v) bromophenol blue
- 0.2 M DTT (Sigma)
- 0.16 M Tris-HCl (pH 6.8)

TBE buffer.
- 90 mM Tris base (BDH)
- 90 mM boric acid (BDH)
- 2 mM EDTA (pH 8.0) (Boehringer Mannheim)

TAE buffer.
- 40 mM Tris base
- 20 mM acetic acid (FSA Laboratory Supplies)
- 1 mM EDTA (pH 8.0)

SDS-PAGE running buffer.
- 0.19 M glycine (BDH)
- 0.1% (w/v) SDS
- 25 mM Tris-HCl (pH 8.3)

Formamide loading buffer.
- 80% (v/v) deionized formamide (BDH)
- 0.01% (w/v) bromophenol blue
- 0.01% (w/v) xylene cyanol
### TABLE 2.2 Conditions for restriction endonuclease digestions

10 x restriction buffers:

<table>
<thead>
<tr>
<th></th>
<th>Medium salt</th>
<th>High salt</th>
<th>Very high salt</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>0.5 M</td>
<td>1 M</td>
<td>1.5 M</td>
</tr>
<tr>
<td>Tris-HCl (pH 7.5)</td>
<td>0.1 M</td>
<td>0.1 M</td>
<td>0.1 M</td>
</tr>
<tr>
<td>Magnesium chloride</td>
<td>0.1 M</td>
<td>0.1 M</td>
<td>0.1 M</td>
</tr>
<tr>
<td>DTT</td>
<td>10 mM</td>
<td>10 mM</td>
<td>10 mM</td>
</tr>
<tr>
<td>BSA</td>
<td>1 mg/ml</td>
<td>1 mg/ml</td>
<td>1 mg/ml</td>
</tr>
</tbody>
</table>

Digestion conditions:

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Manufacturer</th>
<th>Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>BamHI</td>
<td>New England Biolabs</td>
<td>high salt</td>
</tr>
<tr>
<td>BglII</td>
<td>Pharmacia</td>
<td>high salt</td>
</tr>
<tr>
<td>DdeI</td>
<td>Pharmacia</td>
<td>medium salt</td>
</tr>
<tr>
<td>EcoRV</td>
<td>BRL</td>
<td>very high salt</td>
</tr>
<tr>
<td>HindIII</td>
<td>Pharmacia</td>
<td>high salt</td>
</tr>
<tr>
<td>HinfI</td>
<td>New England Biolabs</td>
<td>high salt</td>
</tr>
<tr>
<td>NheI</td>
<td>New England Biolabs</td>
<td>medium salt</td>
</tr>
<tr>
<td>PvuII</td>
<td>BRL</td>
<td>medium salt</td>
</tr>
<tr>
<td>XbaI</td>
<td>Pharmacia</td>
<td>high salt</td>
</tr>
</tbody>
</table>

All incubations were performed at 37°C.
2.3.2. Preparative digestions.

Preparative digestions used up to 200μg of plasmid DNA and 0.5-1 units of enzyme/μg of DNA in a final volume of 200-500μl reaction buffer. Reactions were incubated for 3-12 h and an aliquot analyzed on a mini-gel as described above. If the reaction was complete the reaction was phenol/chloroform extracted and EtOH precipitated. The reaction products were then redissolved in 20-50μl dH₂O and the fragment of interest purified by electrophoresis through a low gelling temperature (LGT; Seaplaque) preparative agarose gel (14cm x 12cm). The correct fragment was visualized using the UV transilluminator and excised with a sterile scalpel blade. The LGT agarose was melted at 65°C for 10 min and NaCl (Sigma) added to a final concentration of 0.3 M. The DNA was then extracted once with phenol equilibrated with 0.3 M NaCl, phenol/chloroform extracted several times with vigorous vortexing until no interface remained, chloroform extracted, and precipitated with 2 volumes 96% (v/v) EtOH at -80°C for 15 min. Finally the DNA was washed in 70% (v/v) EtOH, air dried and resuspended in dH₂O.

2.4 SUBCLONING

2.4.1. Dephosphorylation of DNA.

5' phosphate groups were removed from the vector restriction fragment using calf intestinal alkaline phosphatase (CIP; Boehringer Mannheim). Dephosphorylation of 0.1-1μg of purified DNA restriction fragment was performed in a 50μl volume of 0.5 M Tris-HCl (pH 9.0), 10 mM MgCl₂ (Sigma). The sample received two serial incubations with 0.01 units of CIP at 37°C for 30 min each. The DNA was then phenol/chloroform extracted, precipitated and resuspended in dH₂O.
2.4.2. Ligation.

Ligations were generally performed using DNA in LGT agarose. The agarose was melted at 65°C for 5 min then placed at 37°C for 5 min prior to setting up the reaction at room temperature. A 5:1 molar ratio of insert DNA:dephosphorylated DNA at 100 μg/ml was incubated with 10 units of T4 DNA ligase (New England Biolabs) in 66 mM Tris-HCl (pH 7.6), 10 mM MgCl₂ (BDH), 15 mM DTT (Sigma), 0.2 mg/ml BSA (Sigma) and 1 mM ATP (Pharmacia) for ligations of fragments with over-hangs, or 0.1 mM ATP for blunt-ended ligations, at 14°C overnight.

2.4.3. Preparation of competent cells.

50μl of an overnight culture of the DH5α strain of E. coli (Table 2.3) was used to inoculate 25ml BHI medium (Difco) and grown at 37°C with shaking at 200 rpm until OD₆₀₀ reached 0.3-0.4. This culture was used to inoculate 2 x 250ml BHI at 37°C in 1 litre baffled flasks and again grown at 37°C with shaking at 200 rpm until OD₆₀₀ reached 0.4-0.5. The cells were pelleted by centrifugation at 3000 rpm for 10 min at 0°C in a Sorvall RT6000B centrifuge. The cells were thoroughly drained, resuspended in 100ml of 100 mM rubidium chloride (Sigma), 50 mM manganese chloride (Sigma), 30 mM potassium acetate (BDH), 10 mM calcium chloride (BDH) and 15% (v/v) glycerol (FSA Laboratory Supplies), mixed by inversion and repelleted as before. The cells were resuspended in 10ml 9 mM MOPS-HCl (pH 7.0; Sigma), 10 mM rubidium chloride, 75 mM calcium chloride, 15% (v/v) glycerol and immediately aliquoted to chilled 1ml tubes and stored at -70°C. These competent cells were used for routine transformations.

The MV1190 and CJ236 strains of E. coli (Table 2.3) were used in the site-directed mutagenesis procedure (Bio-Rad Muta-Gene phagemid in vitro mutagenesis kit). These cell types were rendered competent as suggested in the Bio-Rad kit manual. 250ml L-broth (Table 2.4) were inoculated with an overnight culture to an OD₆₀₀ of 0.1 and grown at 37°C with shaking at 200 rpm until OD₆₀₀ reached 0.9. The cells were pelleted at 3000 rpm for 5 min at 0°C, drained and gently
TABLE 2.3 E. coli strains

The following strains of E. coli were used during the work described:-

E. coli CJ236: dut, ung, thi, relA; pCJ105 (Cm'). This strain was maintained in the presence of 30 μg/ml chloramphenicol to prevent loss of the F' plasmid pCJ105.

E. coli MV1190: Δ(lac-proAB), thi, supE, Δ(srt-recA)306::Tn10(tet') [F':traD36, proAB, lacIq Δ M15]. This strain was maintained on glucose-minimal medium plates to select for the F' plasmid which enables growth in the absence of proline.

E. coli DH5α: F'80d-lacZ M15, Δ(lacZYA-argF)U169, recA1, endA1, hsdR17(r−k,m−k), supE44, thi−1, gyrA, relA.

E. coli strains were stored as frozen stocks at -80°C in 30% glycerol, or maintained on suitable agar plates at 4°C.
resuspended in 50ml of ice-cold 100 mM MgCl₂. The cells were then immediately repelleted and gently resuspended in 10ml of ice-cold 100 mM CaCl₂. A further 100ml of 100 mM CaCl₂ were added and the cells held on ice for 1 h. The cells were again pelleted, drained thoroughly and resuspended in 12.5ml of 85 mM CaCl₂, 15% (v/v) glycerol. The cell suspension was immediately aliquoted into chilled 1ml tubes and stored at -70°C.

2.4.4. Transformation.

Ligations were heat treated for 10 min at 65°C, cooled to room temperature and 5μl of ligation reaction added to 200μl of competent E. coli cells (thawed on ice). The cells were incubated on ice for 20 min, heat-shocked at 42°C for 2 min, chilled on ice for 2 min then 800μl of L-broth (Table 2.4) were added and the mixture incubated at 37°C for 60 min with shaking at 150 rpm. The cells were then collected by microcentrifugation at 5,000 rpm for 1 min, and resuspended in 100μl of L-broth. Where appropriate the lacZ gene product was assayed for by the addition of 125μl 4% (w/v) X-gal (Sigma) dissolved in dimethyl formamide (Sigma) and 100μl of 2.4% (w/v) IPTG (Pharmacia) to the mixture. The cells were then plated out on L-agar (Table 2.4) plates supplemented with an appropriate antibiotic for plasmid selection (Table 2.5) and incubated at 37°C overnight. Plates were stored at 4°C.

2.4.5. Identification of positive clones.

Where presence of an insert in the vector disrupted the lacZ gene this was assayed for using X-gal as a substrate and IPTG as an inducer of the gene. Colonies expressing β-galactosidase metabolized X-gal to produce a blue reaction product. blue colonies did not therefore contain an insert and only white colonies were analyzed. 12 colonies were picked from each plate and transferred to 10ml of L-broth supplemented with an appropriate antibiotic for plasmid selection (Table 2.5) in a sterile disposable universal tube.
TABLE 2.4 Media used for growth of *E. coli*

<table>
<thead>
<tr>
<th>Medium</th>
<th>Components</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-broth:</td>
<td>1% (w/v) tryptone (Difco)</td>
<td>0.5% (w/v) Yeast extract (Difco)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5% (w/v) NaCl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>adjusted to pH 7.5 with NaOH</td>
</tr>
<tr>
<td>2 x YT broth:</td>
<td>1.6% (w/v) tryptone</td>
<td>1% (w/v) yeast extract</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5% (w/v) NaCl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>adjusted to pH 7.5 with NaOH</td>
</tr>
<tr>
<td>Glucose-minimal medium:</td>
<td>90 mM sodium phosphate</td>
<td>22 mM potassium phosphate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19 mM ammonium chloride</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5% (w/v) NaCl</td>
</tr>
<tr>
<td></td>
<td>after autoclaving the following were added:</td>
<td>1 mM magnesium sulphate (BDH)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001% (w/v) thiamine-HCl (Sigma)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4% (w/v) glucose (Sigma)</td>
</tr>
<tr>
<td>Agar plates:</td>
<td>To prepare agar plates medium was supplemented</td>
<td>with 1.5% (w/v) Difco agar prior to autoclaving.</td>
</tr>
<tr>
<td>Antibiotics:</td>
<td>Media were supplemented with the following</td>
<td>antibiotics where appropriate:</td>
</tr>
<tr>
<td></td>
<td>Ampicillin: Stock solution is 100 mg/ml of the sodium</td>
<td></td>
</tr>
<tr>
<td></td>
<td>salt (Sigma) in dH₂O, sterilized by filtration and</td>
<td>stored at -20°C. Working solution is 100 mg/ml.</td>
</tr>
<tr>
<td></td>
<td>stored at -20°C. Working solution is 100 mg/ml.</td>
<td>Chloramphenicol: (Sigma) Stock solution is 30 mg/ml in 96% (v/v) EtOH</td>
</tr>
<tr>
<td></td>
<td>Kanamycin: (Sigma) Stock solution is 50 mg/ml in dH₂O,</td>
<td>stored at -20°C. Working solution is 30 μg/ml.</td>
</tr>
<tr>
<td></td>
<td>sterilized by filtration and stored at -20°C.</td>
<td>Kanamycin: (Sigma) Stock solution is 50 mg/ml in dH₂O, sterilized by</td>
</tr>
<tr>
<td></td>
<td>Working solution is 70 μg/ml.</td>
<td>filtration and stored at -20°C.</td>
</tr>
</tbody>
</table>

83
<table>
<thead>
<tr>
<th>Plasmid Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>pRSVcat</strong></td>
<td>The Rous sarcoma virus (RSV) genome from nucleotides -498 to +36 fused to the chloramphenicol acetyltransferase (cat) gene coding sequences and the SV40 polyadenylation signal in a vector background of pBR322. Ampicillin resistant. (Gorman et al., 1982a).</td>
</tr>
<tr>
<td><strong>pMSVcat</strong></td>
<td>The Moloney murine sarcoma virus (Mo-MSV) genome from nucleotides -539 to +30 linked to the cat gene coding sequences and the SV40 polyadenylation signal in a vector background of pBR322. Ampicillin resistant. (Laimins et al., 1984).</td>
</tr>
<tr>
<td><strong>pLTR10cat</strong></td>
<td>Mo-MSV from nucleotides -196 to +30 linked to the cat gene coding sequences and the SV40 polyadenylation signal in a vector background of pBR322. Ampicillin resistant. (Laimins et al., 1984).</td>
</tr>
<tr>
<td><strong>pLTR0cat</strong></td>
<td>Mo-MSV from nucleotides -112 to +30 linked to the cat gene coding sequences and the SV40 polyadenylation signal in a vector background of pBR322. Ampicillin resistant. (Laimins et al., 1984).</td>
</tr>
<tr>
<td><strong>pHPRTcat</strong></td>
<td>The mouse hypoxanthine phosphoribosyltransferase (HPRT) upstream sequences from nucleotides -638 to +48 linked to the cat gene coding sequences and the SV40 polyadenylation signal in a vector background of pBR322. Ampicillin resistant. (A kind gift from D. Stott, NIMR, London).</td>
</tr>
<tr>
<td><strong>pCMVlacZ</strong></td>
<td>The cytomegalovirus immediate early gene (Akrigg et al., 1985) from nucleotides -350 to +72 linked to the lacZ gene coding sequences and the SV40 polyadenylation signal in a vector background of pBR322. Ampicillin resistant. (A kind gift from D. Stott).</td>
</tr>
<tr>
<td><strong>pRSVlacZII</strong></td>
<td>The RSV promoter and enhancer linked to the lacZ gene coding sequences and the SV40 polyadenylation signal in a vector background of pUC8. Ampicillin resistant. (Gorman et al., 1983; Hall et al., 1983).</td>
</tr>
</tbody>
</table>
pRSV_E1A: - The RSV promoter and enhancer sequences linked to the E1A coding sequences in a vector background of pBR322. Ampicillin resistant. (van Dam et al., 1989; a kind gift from N. Jones, ICRF, London).


pT3T718U: - Prokaryotic cloning vector. A phagemid containing the f1 phage ori. Ampicillin resistant. (Pharmacia).


pBluescriptII SK: - Prokaryotic cloning vector derived from pUC19. A phagemid containing the f1 phage ori and the colE1 ori. Ampicillin resistant. (Stratagene).
Cultures were incubated overnight at 37°C with rotation at 150 rpm. The cells were pelleted by centrifugation in a Sorvall RT6000B at 3000 rpm for 5 min, resuspended in 200μl of 25 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0) and then transferred to a microcentrifuge tube. 400μl of 0.2 M NaOH (BDH), 1% (w/v) SDS (BDH) were added to the suspension and gently mixed by inversion. The mixture was neutralized with 300μl of 3 M potassium acetate (pH 4.8), mixed by inversion and microcentrifuged for 10 min at 14,000 rpm. The supernatant was treated with 50μg of DNAse-free RNAse A (Boehringer Mannheim) for 30 min at room temperature, then phenol/chloroform extracted and precipitated. The recovered plasmid was then analyzed by restriction endonuclease digestion (section 2.3.1) or by plasmid sequencing (section 2.7).

2.5 SITE-DIRECTED MUTAGENESIS

The Bio-Rad muta-gene phagemid in vitro mutagenesis kit was used for the generation of point mutations, and the protocols described in the kit manual were used throughout. The kit is based upon a method described by Kunkel (1985; Kunkel et al., 1987) which increases efficiency of mutant generation by selecting against the wild-type strand of dsDNA. A dut, ung double mutant bacterial strain CJ236 (Table 2.3) is used to synthesize DNA containing uracils in thymine positions. The dut mutation inactivates dUTPase resulting in high intracellular concentrations of dUTP, the ung mutation inactivates uracil-N-glycosylase thus allowing the uracil to remain in the DNA. The uracil-containing strand is used as a template for the in vitro synthesis of a complementary strand primed by an oligonucleotide containing the desired mutation. On transformation of the dsDNA into MVI90 cells (Table 2.3) with active uracil-N-glycosylase, the uracil containing strand is inactivated leaving the non-uracil-containing mutated strand to replicate.

2.5.1. Growth of Uracil-containing phagemids.

The sequence to be mutated was subcloned into the phagemid pT3T718U
Double-stranded phagemid was transformed into competent CJ236 cells (see section 2.5.3 and 2.5.4) and plated out on L-agar plates supplemented with 100 μg/ml ampicillin (Sigma; Table 2.4). A single CJ236 colony containing the phagemid was streaked out onto an L-agar plate supplemented with 30 μg/ml chloramphenicol and incubated at 37°C until distinct colonies appeared. An isolated colony was used to inoculate 20ml L-broth supplemented with 30 μg/ml chloramphenicol (Table 2.4) and incubated at 37°C with shaking at 200 rpm overnight. 1ml of this culture was used to inoculate 50ml of 2 x YT media (Table 2.4) containing 100 μg/ml ampicillin and incubation continued until OD₆₀₀ reached 0.3. Helper phage M13K07 was then added to give a multiplicity of infection of approximately 20. Incubation was continued for 1 h then 70μl of 50 mg/ml kanamycin (Sigma) were added and incubation continued for 4-6 h. 30ml of the culture were centrifuged at 12,000 rpm at 4°C for 15 min in a Sorvall SS-34 rotor, using a Sorvall RC5-C centrifuge. The supernatant containing the phagemid particles was transferred to a fresh tube and recentrifuged. 150μg of RNase A were added to the second supernatant and incubated at room temperature for 30 min. 10ml 3.5 M ammonium acetate (Sigma) and 10ml 20% (w/v) PEG 6000 (BDH) were added to the supernatant, and the mixture incubated on ice for 30 min. The phagemids were collected by centrifugation at 12,000 rpm in a Sorvall SS-34 rotor for 15 min and resuspended in 200μl of 300 mM NaCl, 100 mM Tris-HCl (pH 8.0), 1 mM EDTA. The phagemids were then held on ice for 30 min and microfuged for 2 min to remove insoluble material. The phagemids were stored at 4°C and DNA extracted within three days.

2.5.2. Extraction of phagemid DNA.

The 200μl phagemid stock was extracted 2 x with an equal volume of phenol, 1 x with phenol/chloroform and several times with chloroform until there was no visible interface. The aqueous phase was then precipitated with 1/10 volume 7.8 M ammonium acetate and 2.5 volumes 96% (v/v) EtOH at -70°C for 30 min, microfuged for 15 min, washed with 90% (v/v) EtOH and resuspended in 20μl TE. The DNA concentration was determined by electrophoresing a small aliquot on an agarose gel.
with a standard of known concentration.

2.5.3. Synthesis of the mutagenic strand.

The single-stranded uracil-containing DNA prepared above was used as a template for in vitro mutagenesis. The mutagenic oligonucleotide was phosphorylated as described in section 2.8.1, but using non-radioactive ATP. 6-9pmoles of the primer oligonucleotide were then annealed to 200ng of ssDNA in 10μl of 2 mM Tris-HCl (pH 7.4), 0.2 mM MgCl₂, 5 mM NaCl by heating to 70°C and allowing to cool to room temperature over a 45 min period. A control reaction with no primer was performed in parallel. Once the annealing reactions had cooled to room temperature the reactions were placed on ice and the following reagents added: 1μl of 0.4 mM each dNTP, 0.75 mM ATP, 17.5 mM Tris-HCl (pH 7.4), 3.75 mM MgCl₂, 1.5 mM DTT, 3 units of T4 DNA ligase (Bio-Rad) and 1 unit T4 DNA polymerase (Bio-Rad) diluted in ice-cold 100 mM potassium phosphate (pH 7.0), 5 mM DTT, 50% (v/v) glycerol. The reactions were incubated on ice for 5 min then at 25°C for 5 min and finally at 30°C for 90 min. The reactions were then stopped with 90μl of 10 mM Tris-HCl (pH 8.0), 10 mM EDTA and stored at -20°C prior to gel analysis and transformation. 10μl of the reaction products were analyzed by electrophoresis through a 1% (w/v) agarose gel in 1 x TAE buffer (Table 2.1) containing 0.5 μg/ml EtBr. After destaining in dH₂O the reaction products were visualized on the UV transilluminator. In the presence of a primer oligonucleotide, RF-IV covalently closed circular DNA was produced in successful reactions.

2.5.4. Transformation into MV1190 and analysis of transformants.

Competent MV1190 cells were prepared (section 2.4.3.), and transformed with 10μl of the mutagenesis reaction products (section 2.4.4.), and the no primer reaction. 10 and 100μl of the reactions were plated onto L-agar plates supplemented with 100 μg/ml ampicillin. After an overnight incubation colonies were picked and analyzed by sequencing (sections 2.4.5., 2.7).
2.6 PREPARATION OF PLASMID DNA

500ml cultures of plasmid-transformed E. coli were grown overnight at 37°C in baffled 2 litre flasks, with constant shaking at 150 rpm, in BHI medium supplemented with an appropriate antibiotic for selection of the plasmid. Cells were harvested by centrifugation at 5000 rpm, 4°C in a Sorvall RC5-C centrifuge in a GSA rotor. The pellet was resuspended in 40ml of 50 mM glucose (Sigma), Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 5 mg/ml lysozyme (Sigma), and incubated on ice for 10 min. 80ml of 0.2 M NaOH, 1% (w/v) SDS were added, mixed gently by inversion and the mixture incubated on ice for 10 min. 40ml of 3 M potassium acetate (pH 4.8) were then added to neutralize the mixture and incubation continued for a further 10 min on ice. The mixture was then centrifuged at 10,000 rpm for 15 min at 4°C in the GSA rotor and the supernatant filtered through muslin. 0.6 volumes of propan-2-ol (BDH) were added and plasmid DNA precipitated for 20 min at room temperature. The precipitate was collected by centrifugation at 10,000 rpm for 15 min at 4°C in the GSA rotor. The pellet was resuspended in 9ml of TE (10 mM Tris-HCl (pH 8.0), 1 mM EDTA) and neutralized by the addition of 1ml of 1 M Tris base. 10.2g of caesium chloride (FSA Laboratory Supplies) were dissolved in the solution and EtBr added to a final concentration of 1 mg/ml. The density of the solution was determined by weighing and adjusted, if necessary, to 1.55 mg/ml. The solution was then transferred to a Beckman Ti 70.1 polyallomer quick-seal tube and centrifuged for 24 h at 47.5 Krpm, 20°C in a Ti 70.1 rotor in a Beckman L8-70M ultracentrifuge. Supercoiled plasmid DNA was collected using a syringe and 19 G needle by side-puncture of the tube under long-wave UV illumination. The EtBr was removed by repeated extractions to dH₂O saturated butan-1-ol, and the CsCl removed by extensive dialysis against 2 x 1 litre of TE at 4°C with stirring. The DNA was then EtOH precipitated and redissolved in dH₂O. Plasmid preparations were stored at 4°C.

2.7 SEQUENCING

Dideoxy sequencing was performed using Sequenase enzyme (U.S.
Biochemical Corporation) in accordance with the manufacturer's instructions. Plasmid DNA was prepared for sequencing either by using CsCl/EtBr isopycnic density gradients (section 2.6) or by small-scale alkaline lysis preparations (section 2.4.5.) followed by a second precipitation step in 6.5% (w/v) PEG 6000, 0.4 M NaCl, and thorough washing with 70% (v/v) EtOH.

5µg of DNA to be sequenced were alkaline denatured in 25µl of 200 mM NaOH, 0.5 mM EDTA for 5 min at room temperature. The mixture was then precipitated with 10µl of 5 M ammonium acetate (pH 7.5) and 70µl of EtOH at -80°C for 15 min, and washed with 70% (v/v) EtOH. To the DNA pellet the following were added - 7µl dH₂O, 2µl sequenase buffer (Table 2.6), 0.5pM primer oligonucleotide, 1µl 100 mM DTT, 2µl of a 1:5 dilution labelling mix : dH₂O (Table 2.6), 5µCi [$\alpha$-35S] dATP (Amersham) and 2µl Sequenase enzyme (1:8 dilution in TE (pH 7.5)). After 10 min at room temperature 3.5µl of each reaction mix were added to 2.5µl of the 4 termination mixes (Table 2.6), in a 60 well micro-titre plate. The tubes were incubated at 37°C for 30 min before addition of 4µl of sequenase loading buffer (Table 2.6). Samples were boiled for 2 min and chilled on ice prior to electrophoresis.

Sequencing gels were cast between clean glass plates of 38cm by 20cm, one of which was siliconized with "repelcote" (BDH), with 0.4mm spacers and a "sharks tooth" comb. 400µl of 10% (w/v) ammonium persulphate (BDH) and 40µl TEMED (Sigma) were added to 40ml of 7% (w/v) acrylamide (BDH), 0.3% (w/v) bis-acrylamide (BDH), 7M urea (Sigma), 1 x TBE (Table 2.1). The solutions were mixed thoroughly and the gel cast immediately. Electrophoresis was performed in 1 x TBE at 40 W constant power for 2-4 h. Gels were pre-electrophoresed for 45 min prior to loading the samples. After electrophoresis the gel was fixed in 10% (v/v) acetic acid (FSA Laboratory Supplies), 10% (v/v) methanol (Hayman Ltd.) for 20 min at room temperature, then transferred to Whatmann 3MM paper, overlaid with Saran wrap and dried under vacuum at 80°C. Once the gel was dry it was exposed to Kodak X-ray (XAR-5) film, at room temperature.
Table 2.6 SEQUENCING BUFFERS

Sequenase buffer.
- 200 mM Tris-HCl (pH 7.5)
- 100 mM MgCl₂
- 250 mM NaCl

Labelling mix.
- 7.5 μM dGTP
- 7.5 μM dCTP
- 7.5 μM dTTP

ddG Termination mix.
- 80 μM dGTP
- 80 μM dATP
- 80 μM dCTP
- 80 μM dTTP
- 8 μM ddGTP
- 50 mM NaCl

ddA Termination mix.
As for ddG Termination mix except that ddGTP is replaced by ddATP.

ddT Termination mix.
As for ddG Termination mix except that ddGTP is replaced by ddTTP.

ddC Termination mix.
As for ddG Termination mix except that ddGTP is replaced by ddCTP.

Sequenase gel loading buffer.
- 95% (w/v) deionized formamide
- 20 mM EDTA
- 0.05% (w/v) xylene cyanol
- 0.05% (w/v) bromophenol blue
2.8 END LABELLING

2.8.1. Labelling using T4 polynucleotide kinase.

Oligonucleotides were 5'-end labelled using T4 polynucleotide kinase (BRL). 250ng of oligonucleotide were labelled in 50 mM Tris-HCl (pH 7.6), 10 mM MgCl₂, 5 mM DTT, 0.1 mM spermidine (Sigma), 0.1 mM EDTA with 100μCi [γ-³²P] ATP (NEN) and 10 units of T4 polynucleotide kinase for 45 min at 37°C. A second aliquot of enzyme was then added and the reaction continued for a further 45 min before phenol/chloroform extraction and ethanol precipitation. The precipitate was washed several times with 70% (v/v) EtOH to remove unincorporated ³²P.

2.8.2. Labelling using DNA polymerase I (Klenow fragment).

Restriction fragments were 3'-end labelled using DNA polymerase I (Klenow fragment) (Pharmacia). 200-500ng of restriction fragment were incubated with 20-50μCi of [α-³²P] dNTP (Amersham) and 3 units of Klenow enzyme in 20μl of medium salt restriction endonuclease buffer (Table 2.2) for 20 min at room temperature. The dNTP was chosen to be complementary to the first nucleotide of the 5' overhang. The reaction was then phenol/chloroform extracted and EtOH precipitated, and the precipitate washed several times with 70% (v/v) EtOH to remove unincorporated ³²P.

2.8.3. Measurement of incorporation of ³²P.

The percentage incorporation of ³²P was measured by TCA precipitation and scintillation counting. 1μl of labelled probe was added to 0.5ml salmon sperm DNA (Sigma) at 500 μg/ml in 20 mM EDTA and 5ml ice-cold 10% (w/v) TCA (Sigma). After incubation on ice for 15 min the mixture was filtered through a Whatman GF/C filter paper disk. The disk was washed with 6 x 5ml 10% (w/v) TCA then with 96% (v/v) EtOH, air-dried and transferred to a scintillation vial. 5ml of Econofluor scintillant (NEN) were added and ³²P-incorporation...
determined by scintillation counting in a Beckman LS 6000IC
scintillation counter.

2.9 TISSUE CULTURE

All cells were maintained at 37°C, 8.5% (v/v) CO₂, 90% humidity on
Nunclon tissue culture plasticware (Gibco-BRL). All cell lines were
routinely tested for mycoplasma contamination every 4 weeks. Cells
were stored as frozen stocks at 5 x 10⁶ cells/ml in medium plus 10%
(v/v) DMSO (Sigma), in cryotubes under liquid nitrogen. Cells were
thawed rapidly at 37°C and immediately diluted with appropriate
medium before pelleting by centrifugation in an MSE benchtop
centrifuge for 5 min at 1,500 rpm, then resuspended and plated out.

F9 EC cells (Bernstine et al., 1973) were grown in Dulbecco’s
modification of Eagle’s medium (DMEM; Flow) supplemented with 10%
(v/v) foetal calf serum (FCS; Imperial Laboratories), 500 units/ml
penicillin (Sigma) and 100 μg/ml streptomycin sulphate (Sigma).
Subconfluent cultures were passaged every 3 days by washing with 10mL
PBS, and trypsinizing with 1ml trypsin-versene (for 1 litre:- NaCl
8g, KCl 0.2g, Na₂HPO₄ 1.15g (Sigma), KH₂PO₄ 0.2g (Sigma), EDTA 0.1g,
Trypsin 1.25g (Sigma), Phenol Red 0.01g (Sigma) in dH₂O (pH 7.5)
stored at -20°C) for 2 min at room temperature. Once the monolayer
had detached from the dish the reaction was stopped by the addition
of medium. Due to the tendency of F9 EC cells to clump they were
passed through a 19 G syringe needle prior to harvesting by
centrifugation for 5 min at 1,500 rpm in an MSE benchtop centrifuge.
The cells were resuspended at 2 x 10⁴/ml and plated out on Nunclon
dishes.

The PCC4 EC cell line (Jakob et al., 1973) was treated in a similar
fashion, as was the SV3T3 clone 49 murine fibroblast line (Rigby et
al., 1980) except that these cells were not passed through a 19 G
needle. HeLa cells (a transformed epithelial cell line derived from a
human cervical carcinoma, Gey et al., 1952) were maintained in DMEM
supplemented with newborn calf serum (NCS) rather than FCS.
2.10 F9 CELL DIFFERENTIATION IN VITRO

F9 EC cells were differentiated to parietal endoderm-like cells by the method of Strickland et al. (1980). After passaging the cells were plated out at a density of $10^5$/ml in medium supplemented with $10^{-7}$ M all-trans retinoic acid (made up as a 10 mM stock in EtOH; Sigma) and $10^{-4}$ M dibutyryl cAMP (Sigma). After 48 h the cells were passaged in the same manner. After a further 60 h differentiation was judged to be complete by examination of morphology and loss of the antigen SSEA-1.

2.11 ANTI-SSEA-1 STAINING

Presence of the stage-specific embryonic antigen (SSEA-1) was detected using the monoclonal antibody MC 480 (Solter and Knowles, 1981; a kind gift from D. Solter). 90 mm diameter plates of cells were washed twice with 10 ml PBS and then fixed with 10 ml 4% (v/v) formalin (Sigma) in PBS for 15 min at room temperature. The cells were then rendered permeable by incubation with 10 ml 0.1% (v/v) Triton X100 (Sigma) in PBS for 15 min at room temperature. After two 10 ml washes in PBS, anti-SSEA-1 antibody was added as 10 ml of a 1:1000 dilution of ascites fluid in PBS, 1% (v/v) FCS and incubated for 30 min at 37°C. Control reactions with no antibody were also performed.

Unbound antibody was removed with four 10 ml washes of PBS, 1% (v/v) FCS. 5 ml of a 1:100 dilution of rabbit anti-mouse immunoglobulins conjugated to peroxidase (Dako Immunoglobulins) in PBS were then added and incubation continued at 37°C for 30 min. The cells were washed a further four times in PBS before locating antibody-bound cells using a substrate for the peroxidase enzyme consisting of 0.1% (w/v) dianisidine (Sigma; from a saturated solution in EtOH), 0.1% (v/v) hydrogen peroxide (BDH) in PBS. The reaction was performed for 20 min at 37°C in the dark. The cells were then rinsed thoroughly in PBS prior to observation by microscopy.
2.12 TRANSFECTION

Transfections were performed essentially as described by Gorman et al. (1982b). Supercoiled plasmid DNA for transfection was purified twice by CsCl/EtBr isopycnic density gradient centrifugation. 5 x 10⁶ cells were transfected 5 h after plating out as single-cell suspensions in 50mm diameter dishes in medium buffered with 10 mM Hepes-HCl (Sigma tissue culture reagents) to pH 7.3. Precipitates were formed by mixing 5μg or 10μg total DNA (held constant within each experiment) comprising a test construct plus plasmid carrier DNA in 250μl of 250 mM CaCl₂, TE (pH 7.12) with 250μl of 0.3 M NaCl, 3 mM Na₂HPO₄ (Sigma), 50 mM Hepes-HCl (pH 7.12) dropwise over 2 min. The precipitates were then added dropwise over the surface of the monolayer of cells. After an 18 h incubation the cells were washed with 2 x 5ml of DMEM at 37°C and then refed with 5ml of complete medium. After a further 30 h incubation the cells were washed with 5ml PBS, and scraped into 0.5ml of 0.15 M NaCl, 1 mM EDTA, 40 mM Tris-HCl (pH 7.5) using a sterile cell-scaper and transferred to a microcentrifuge tube. The cells were harvested by microcentrifugation for 1 min and aspirated thoroughly prior to extraction of protein. Transfections were generally performed in triplicate.

2.13 CAT ASSAYS

Extracts of protein for chloramphenicol acetyltransferase (CAT) assay from transfected cells were performed in 100μl of 100 mM Tris-HCl (pH 7.8) by rapid freezing and thawing (-70°C to 37°C) three times. The cell debris was then removed by microcentrifugation for 5 min and the extract transferred to a clean tube. Contaminating acylases (Crabb and Dixon, 1987) were removed by heat treatment at 65°C for 10 min, and the debris was again removed by microcentrifugation. 5μl of the extract were added to 1ml of 1/10 dilution of Bio-Rad protein concentration assay and the protein concentration determined relative to BSA.

CAT assays were performed by the "one-vial" quantitative method (Sleigh, 1986). Reactions were set up on ice in 7ml plastic
scintillation vial inserts. 100μg of protein extract in 50μl of Tris-HCl (pH 7.8) were added to 200μl of 1.25 mM chloramphenicol (Sigma), 1% (v/v) EtOH, 100 mM Tris-HCl (pH 7.8) and 0.1μCi [14C] butyryl CoA (NEN) at 1 min intervals and overlayed with 5ml Econofluor scintillant (NEN). The reactions were allowed to proceed at room temperature, and the reaction products, which have a greater solubility in organic solutions than the reagents, assayed by scintillation counting at 1 min intervals. Scintillation counting was repeated several times, this allowed a reaction rate to be calculated which reflected the concentration of CAT enzyme in the extract. This value allows the activity of the promoter driving expression of the CAT coding sequences to be determined.

2.14 ASSAY FOR β-GALACTOSIDASE PRODUCTION

The production of β-galactosidase enzyme by transfected tissue culture cells was assayed for using X-gal as a substrate. The cells were fixed in 0.5% (v/v) glutaraldehyde (Sigma) in PBS for 15 min at room temperature, and then washed three times in 2 mM MgCl₂ in PBS for 5 min. To each 50mm plate 3ml of 1 mg/ml X-gal in 30 mM K₃Fe(CN)₆ (BDH), 30 mM K₄Fe(CN)₆.3H₂O (BDH), 2 mM MgCl₂, 0.01% (w/v) sodium deoxycholate (Sigma), 0.02% (v/v) NP40 (BDH) were added. The plates were incubated at 37°C in the dark for 3-18 h, washed in PBS and blue staining observed under a microscope.

2.15 RNA EXTRACTION

Total cellular RNA was extracted from tissue-culture cells by the method of Chirgwin et al. (1979). Cell monolayers were washed with PBS, then lysed with 8.5ml of 4 M guanidinium thiocyanate (Fluka; deionized with Bio-Rad AG 501-X8(D) resin) in 25 mM sodium citrate (BDH), 1% (w/v) Sarkosyl (Sigma), 0.1 M β-mercaptoethanol (Sigma), per 5 x 10⁸ cells. The lysate was gathered using a sterile cell-scaper and layered onto 3.5ml of 5.7 M CsCl, 0.1 M EDTA (pH 5.0) in a Beckman polyallomer SW40 tube and centrifuged at 30 Krpm at room temperature for 20 h in an SW40 rotor using a Beckman L8-70M
centrifuge. After centrifugation the supernatant was aspirated and the tube inverted and wiped dry with sterile tissue. The RNA pellet was resuspended in 200μl DEPC treated water, EtOH precipitated, and stored at -70°C in DEPC treated water.

2.16 WHOLE-CELL EXTRACTS

Whole-cell extracts were prepared according to the method of Manley et al. (1980). 30-100 (140mm diameter) plates of sub-confluent cells were harvested in chilled PBS using a sterile cell-scaper. The cells were pelleted by centrifugation at 1,500 rpm for 5 min at 4°C in a Sorvall RT6000B centrifuge and resuspended in 4 packed-cell volumes of 10 mM Tris-HCl (pH 7.9), 1 mM EDTA, 5 mM DTT, 0.1 mM PMSF (Sigma). All subsequent steps were carried out at 4°C. After 20 min on ice the cells were lysed with 8 strokes of a Dounce homogenizer. 4 packed-cell volumes of 50 mM Tris-HCl (pH 7.9), 10 mM MgCl₂, 2 mM DTT, 25% (w/v) sucrose (BDH), 50% (v/v) glycerol were added and the mixture gently stirred. 1 packed-cell volume of a saturated solution of ammonium sulphate (BDH; pH 7.9) was added dropwise to the mixture and gently stirred for 30 min. The cell debris was pelleted by centrifugation at 40 Krpm for 2 h at 4°C in an SW40 rotor using a Beckman L8-70M centrifuge. 0.33 g/ml of solid ammonium sulphate were added to the supernatant and dissolved by stirring for 20 min. 0.33 μl/ml of 1 M NaOH were then added and stirring continued for a further 30 min. The extract was pellet at 28 Krpm for 45 min at 4°C in an SW28 rotor using a Beckman L8-70M centrifuge, and then resuspended in a minimum volume of Manley dialysis buffer (20 mM Hepes-HCl (pH 7.9), 100 mM KCl, 12.5 mM MgCl₂, 2 mM DTT, 0.1 mM EDTA, 17% (v/v) glycerol). Dialysis against 2 x 1000 volumes of Manley dialysis buffer was performed for 16 h, and then insolubles removed by centrifugation at 3,000 rpm for 30 min in a Sorvall RT6000B centrifuge. The extract was aliquoted into 1ml volumes and stored in liquid nitrogen (long-term), or into 10-100μl volumes and stored at -70°C. The concentration of each extract was determined relative to BSA using the Bio-Rad protein concentration assay.
2.17 IN VITRO TRANSCRIPTION

In vitro transcriptions were performed essentially as described by La Thangue and Rigby (1987). Transcription was performed in a volume of 10 μl at a final concentration of 12 mM Heps-HCl (pH 7.9), 60 mM KCl, 7.2 mM MgCl₂, 0.14 mM EDTA, 1.2 mM DTT, 10.2% (v/v) glycerol, 0.5 mM each rNTP. Reactions contained 0.125-2 μg of template DNA and 5-100 μg whole-cell extract. After 40 min incubation at 30°C the DNA was removed by incubation for 30 min at 30°C with 50 μl of 1.5 units/μl RNase-free DNase (Gibco-BRL), 0.45 units/μl RNase-inhibitor (RNA Guard; Pharmacia), 40 mM NaCl, 10 mM MgCl₂, 10 mM Tris-HCl (pH 7.5), 2 mM CaCl₂. The reaction was then phenol/chloroform extracted and EtOH precipitated.

2.18 PRIMER EXTENSION

The products of in vitro transcription reactions and cellular RNA were analyzed by primer extension from an end-labelled oligonucleotide primer complementary to nucleotides +15 to +34 of the coding sequence of the cat gene (see section 2.2). In vitro transcription products were annealed to 2.5 ng of primer in a 3 μl volume of dH₂O at 65°C for 15 min, the primer extension reaction was then carried out in a 10 μl volume as described below. 20 μg of cellular RNA and 0.25 ng primer were boiled for 3 min in 400 mM NaCl, 40 mM Pipes (Sigma; pH 6.5), 0.5 mM EDTA then annealed overnight at an empirically determined temperature between 45°C and 60°C, the primer extension reaction was then carried out in a 50 μl volume. Primer extension reactions were performed at 43°C in a final concentration of 50 mM Tris-HCl (pH 8.3), 50 mM KCl, 10 mM MgCl₂, 10 mM DTT, 0.5 mM dNTPs using 10 units of reverse transcriptase (Seikagaku). Cellular RNA was then removed by incubation for 15 min at 37°C with 1 μg of DNase-free RNase and EDTA to a final concentration of 10 μM. The reactions were then phenol/chloroform extracted and EtOH precipitated before resuspending in formamide loading buffer (Table 2.1), boiling for 3 min and loading onto an 8% (w/v) acrylamide gel. Primer extension gels were cast between clean
glass plates of 38cm by 20cm with 0.4mm spacers and a 0.5cm well-width comb. 400µl of 10% (w/v) ammonium persulphate and 40µl TEMED were added to 40ml of 8% (w/v) acrylamide, 0.3% (w/v) bis-acrylamide, 7 M urea, 1 x TBE (Table 2.1). The solutions were mixed and the gel cast immediately. Once set the gel was pre-electrophoresed for 45 min in 1 x TBE, prior to loading. Molecular weight markers of pBR322 DNA digested with Hinfl endonuclease and ^32P end-labelled were electrophoresed in parallel with the samples. Electrophoresis was continued for approximately 2 h at 40 W, the gel dried and exposed to Kodak XAR X-ray film with intensifying screens at -70°C.

2.19 RNASE PROTECTION

Riboprobes were synthesized from linearized template DNA using T7 RNA polymerase (Stratagene). Reactions consisted of 250ng of linearized template DNA, 0.4 mM rATP, 0.4 mM rCTP, 0.4 mM rGTP, 0.02 mM UTP, 100 µCi [α^32P-UTP], and 5 units T7 RNA polymerase in 200 mM Tris-HCl (pH 8.0), 40 mM MgCl₂, 10 mM spermidine (Sigma), 250 mM NaCl. Reactions were performed for 1 h at 30°C then treated with 2 units of DNaseI (Gibco-BRL) in 10 mM Tris-HCl 7.5, 10 mM MgCl₂ in the presence of 50 units of RNAse-inhibitor (RNA Guard; Pharmacia). The probe was then phenol/chloroform extracted, ethanol precipitated, and resuspended in 10µl of DEPC treated water. 1µl of the probe was electrophoresed through an 8% acrylamide gel for 2 h and exposed to Kodak X-AR film for 30 min (as described in section 2.18) to check that the majority of the probe was full length. 1µl of the probe was TCA precipitated, as described above and added to 5ml Econofluor scintillant (NEN) and the specific activity determined by scintillation counting. The probe was stored at -20°C under ethanol and used within 3 days.

5 x 10⁶ cpm of the riboprobe were coprecipitated with 50µg of RNA (prepared as described in section 2.15). Hybridization was then performed in 40 mM Pipes-HCl (pH 6.4), 400 mM NaCl, 1 mM EDTA, 80% (v/v) formamide by heating at 90°C for 5 min then transferring to 50°C and incubating overnight. The tubes were then placed on ice and RNase digestion performed with 2 µg/ml RNAse T1 (Boehringer Mannheim) and 40 µg/ml RNAse A (Boehringer Mannheim) in 300µl of 0.3 M NaCl, 10
mM Tris-HCl (pH 7.5), 5 mM EDTA for 1 h at 30°C. The reactions were
Proteinase K treated with 100 μg of Proteinase K (Boehringer Mannheim)
in 0.5% (w/v) SDS for 15 min at 37°C, phenol/chloroform extracted and
precipitated, resuspended in 5 μl formamide loading buffer (Table 2.1)
and electrophoresed through an 8% acrylamide gel (as described for
primer extension reactions, section 2.15).

2.20 GEL RETARDATION

Binding reactions for gel-retardation assays (Fried and Crothers,
1981; Garner and Revzin, 1981) were performed in 20 μl volumes under
equivalent conditions to in vitro transcription reactions (section
2.16). 1-10 μg of whole-cell extract were incubated with the stated
concentration of non-specific competitor DNA (generally sonicated
salmon sperm DNA) and specific competitor for 15 min at 30°C. 0.5ng
of [32P] labelled probe (DNA fragment or oligonucleotide) were then
added and incubation continued for a further 15 min. The reactions
were chilled on ice and 10 μl loaded onto a prerun non-denaturing
acrylamide gel and electrophoresed in 1 x TAE (Table 2.1) circulating
buffer at 150 V for 90 min at 4°C. Non-denaturing gels were prepared
by adding 400 μl of 10% (w/v) ammonium persulphate and 40 μl of TEMED
to 40 ml of 4% (w/v) acrylamide, 0.3% (w/v) bis-acrylamide in 1 x TAE,
and cast between clean glass plates of 19.5 cm x 16 cm with 1.5 mm
spacers. Gels were prerun for at least 3 h at 4°C or until the
current stabilized. After electrophoresis gels were dried and exposed
to Fuji RX X-ray film with intensifying screens at -70°C.

A variation of the gel retardation technique was to use antibodies
to further retard complexes which included a protein recognized by
that antibody. In these experiments antibody or pre-immune serum was
added to the reaction either before or after addition of the probe,
as stated. In experiments where purified protein was used, rather
than whole-cell extract, 1 mg/ml BSA was included in the binding
reaction to prevent loss of protein by adsorption to plasticware, and
0.1% (v/v) NP40 was included in the gel and buffer to prevent
aggregate formation.

100
2.2.1 WESTERN BLOT ANALYSIS

Proteins were electrophoresed through a 1.5mm thickness, 7.5% (w/v) SDS-polyacrylamide gel in a Mini Protean gel apparatus (Bio-Rad) essentially according to the method of Laemmli (1970). The resolving gel comprised 7.5% (w/v) acrylamide, 0.15% (w/v) bis-acrylamide, 0.1% (w/v) SDS, 0.37 M Tris-HCl (pH 8.8), and was polymerized by the addition of 1/100 volumes 10% (w/v) ammonium persulphate and 1/1000 volumes TEMED. The resolving gel was cast to 1cm below the bottom of the gel comb and once set was overlaid with the stacking gel. The stacking gel comprised 5% (w/v) acrylamide, 0.1% (w/v) bis-acrylamide, 0.1% (w/v) SDS, and 0.125 M Tris-HCl (pH 8.0), and was polymerized as above.

Samples were boiled for 3 min in SDS-PAGE (SDS-polyacrylamide gel electrophoresis) loading buffer prior to electrophoresis at 200 V for approximately 30 min, or until the bromophenol blue dye reached the bottom of the gel. Bio-Rad prestained low range molecular weight standards were electrophoresed in parallel as molecular weight markers. Transfer of the protein to nitrocellulose was achieved using a semi-dry electroblotting apparatus (Bio-Rad). 6 layers of filter paper soaked in 0.3 M Tris-HCl (pH 10.4), 20% (v/v) methanol were laid on the positive electrode, followed by 3 layers of filter paper soaked in 25 mM Tris-HCl (pH 10.4), 20% (v/v) methanol. These were followed by the nitrocellulose filter, the gel, then 9 layers of filter paper soaked in 25 mM Tris-HCl (pH 9.4), 40 mM 6-amino-n-hexanoic acid (Sigma), 20% (v/v) methanol. A current of 1.5 mA/cm² was applied for 1 h, and the efficiency of blotting assessed by staining any protein remaining in the gel with 0.01% (w/v) Coomassie brilliant blue (BDH) in 10% (v/v) acetic acid, 10% (v/v) methanol.

Blotted antigens were detected using a Promega Protoblot kit, essentially as recommended by the manufacturer. The nitrocellulose filter was removed from the blotting apparatus and washed in TBST (10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.05% (v/v) Tween 20 (Promega)). Excess protein binding sites were then blocked by incubation for 1 h at room temperature in 5% (w/v) BSA (blot qualified; Promega) in TBST. The blocking solution was replaced with TBST containing a
1/1000 dilution of the primary antibody, and binding performed at room temperature for 2 h with constant agitation. Unbound antibody was then removed with three 10 min washes in TBST. The filter was transferred to a 2% (w/v) BSA solution in TBST containing a 1/5000 dilution of anti-IgG-alkaline phosphatase conjugate (Promega), and incubated for 1 h at room temperature with constant agitation. After 3 further washes for 10 min each in TBST the filter was gently blotted dry and transferred to the colour development solution. The colour development solution was prepared by the addition of 66μl of nitro blue tetrazolium (Promega) at 50 mg/ml in 70% (v/v) dimethylformamide to 10ml of 100 mM Tris-HCl (pH 9.5), 100 mM NaCl, 5 mM MgCl₂. After thorough mixing a further addition of 33μl of 5-bromo-4-chloro-3-indolyl phosphate (Promega) at 50 mg/ml in dimethylformamide was made and the solution thoroughly mixed and used immediately. After 2-10 mins the colour development reaction was stopped by rinsing the filter in dH₂O.

2.22 HEPARIN-SEPHAROSE COLUMN CHROMATOGRAPHY

Whole-cell extracts were fractionated on heparin-Sepharose (Pharmacia) columns. Heparin-Sepharose was prepared by soaking in column buffer (20 mM Hepes-HCl (pH 7.9), 20% (v/v) glycerol, 0.5 mM PMSF, 0.5 mM DTT, 12.5 mM MgCl₂) plus 0.1 M KCl overnight, then packed into Bio-Rad Econocolumns and washed with 5 column volumes of column buffer plus 0.1 M KCl at 10 ml/h. All operations were performed at 4°C. Extract was absorbed onto the column and washed with 2 column volumes of column buffer plus 0.1 M KCl at 10 ml/h, followed by a linear gradient, of approximately 5 mM/min of 0.1 M KCl to 1 M KCl in column buffer. Finally the column was washed with 2 column volumes of 1 M KCl in column buffer. 1.3ml fractions were collected at 9 min intervals and stored at -70°C.

2.23 WHEAT GERM AGGLUTININ COLUMN CHROMATOGRAPHY

Columns of the lectin wheat germ agglutinin (WGA) coupled to agarose (Vector Laboratories) were used to absorb glycosylated
proteins. WGA resin was soaked in WGA buffer (20 mM Hepes-HCl (pH 7.9), 20% glycerol, 12.5 mM MgCl₂, 0.5 mM DTT, 0.5 mM PMSF, 0.1% (v/v) NP40, 10 μM zinc sulphate (Sigma), 0.1 M KCl), packed into columns and washed with 10 column volumes of WGA buffer at 10 ml/h. All operations were carried out at 4°C. Whole-cell extract was absorbed onto the column and washed with 5 column volumes of WGA buffer at 10 ml/h. N-acetyl glucosamine (GlcNAc; Sigma) in WGA column buffer was used to elute the protein. Fractions were collected and stored at -70°C.

2.24 AFFINITY CHROMATOGRAPHY

Affinity columns were prepared and used essentially as previously described (Kadonaga and Tjian, 1986). 250μg of plus and minus strand complementary oligonucleotides were phosphorylated using T4 polynucleotide kinase as described in section 2.8.1. but with unlabelled rATP. The complementary strands were heated to 90°C and annealed by cooling to room temperature overnight. Concatamers were then formed by ligation in a 100μl volume under the conditions described in section 2.4.2. The success of the concatamerization reaction was assessed by electrophoresis through a 1% (w/v) agarose gel with size markers of pBR322 DNA digested with Hinfl.

3g of cyanogen bromide activated Sepharose (Pharmacia) were placed in a sintered glass funnel and washed with 500ml 1 mM HCl at 4°C, followed by 500ml 1 mM potassium phosphate (pH 8.0) at 4°C. The resin was transferred to a 50ml disposable polypropylene tube, allowed to settle, and the 1 mM potassium phosphate aspirated. 500μg of concatamerized double-stranded oligonucleotide in 10 mM potassium phosphate (pH 8.0) were then added to a total volume of 40ml. The tube was rotated overnight at 20°C to allow coupling to occur.

The resin was then returned to the sintered glass funnel and washed twice with 100ml dH₂O and once with 100ml 1 M ethanolamine-HCl (pH 8.0; Sigma). The resin was then resuspended in 20ml 1 M ethanolamine-HCl (pH 8.0) in a 50ml tube and rotated for 5h at 20°C. After transfer to a sintered glass funnel the resin was washed with the following in turn: 100ml of 10 mM potassium phosphate (pH 8.0),
100ml of 1 M potassium phosphate (pH 8.0), 100ml of 1 M KCl, 100ml of dH₂O, 100ml of 10 mM Tris-HCl (pH 7.6), 0.3 M NaCl, 1 mM EDTA. The resin was then resuspended in the final wash solution and stored at 4°C.

5ml of oligonucleotide coupled Sepharose were packed in a Bio-Rad Econocolumn and washed with 10 column volumes of heparin-Sepharose column buffer plus 0.1 M KCl (section 2.22). Partially purified extract was then applied to the column which was otherwise used in precisely the same manner as heparin-Sepharose columns.

2.25 COMPUTING

Sequence data were analyzed using the University of Wisconsin Genetics Computer Group’s sequence analysis software package (Devereux et al., 1984) Sun OS version 5.3. CAT assay data were analyzed using the programme "Cricket Graph" version 3.0 on a Mackintosh MS2 micro-computer.
CHAPTER 3

A COMPARISON OF RETROVIRUS LTR-BINDING PROTEINS IN EC CELLS AND THEIR DIFFERENTIATED DERIVATIVES.
3.1 INTRODUCTION

As discussed in section 1.6, the Moloney murine sarcoma virus (Mo-MSV) was chosen as a viral probe with which to investigate transcriptional control in early embryonic stem cells. Mo-MSV was an ideal candidate for this study for the following reasons:-

a) It had been cloned and sequenced.
b) It was readily available.
c) Cis-acting sequence elements responsible for transcription in differentiated cell types had previously been investigated.
d) The sequences responsible for the block to transcription in embryonic stem cells had already been investigated on a gross level.
e) Mutations had been characterized which relieve the block to transcription in a variety of EC cell types.

Due to the practical difficulties of working with actual embryos, e.g. their small size, difficulty of isolation, and heterogeneity of cell type, F9 embryonal carcinoma (EC) stem cells were chosen as a model system to represent early embryonic cells. ES cells might currently be considered the model system of choice. However, at the start of these investigations ES cell technology was in its infancy, the differentiation-inhibiting-activity which obviates the need for feeder cells was unavailable, and transfection of the cells had not been perfected. Conversely, EC cells were known to be simple to grow and to differentiate in vitro. They yield sufficient quantities of protein to perform in vitro experiments, and are amenable to transfection.

The gel retardation technique (Fried and Crothers, 1981; Garner and Revzin, 1981) was chosen to identify sequence elements and DNA binding proteins with a potential role in the control of Mo-MSV transcription in EC cells. This is a sensitive technique which detects protein-DNA complexes by virtue of their retarded migration, with respect to free DNA, through a non-denaturing acrylamide gel. The probes used in gel retardation experiments were DNA fragments from the wild-type Mo-MSV LTR, or synthetic oligonucleotides representing wild-type Mo-MSV sequences or the corresponding
sequences from the mutant viruses MPSV and PCMV. Whole-cell extracts (Manley et al., 1980) were made from F9 EC cells, PCC4 EC cells, their differentiated derivatives, and fibroblastic cell lines for use in these assays. By comparing the binding patterns between wild-type and mutant probes, and between undifferentiated and differentiated cell extracts, sequences and proteins with a potential role in control of transcription were identified for further analysis.

The major comparison was between F9 EC cells and parietal endoderm-like cells, the differentiated derivatives of F9 cells. It was predicted that EC cells might either contain a repressor of transcription responsible for the EC cell block to Mo-MSV transcription, or lack an activator of transcription that is present in their differentiated derivatives. It was therefore important that extracts were made from cells showing the true undifferentiated or the fully differentiated phenotype. Phenotype was monitored by probing for the stage-specific early antigen-1 (SSEA-1; Solter and Knowles, 1981), which is no longer expressed following differentiation of EC cells, and by observation of cell morphology. Fig. 3.1 shows typical samples of F9 EC cells and parietal endoderm-like cells fixed and stained with anti-SSEA-1 antibody. The F9 EC cells displayed the characteristic morphology of tightly clumped cells in rounded colonies and they stained brown indicating that they were expressing SSEA-1. However, the majority of the F9 cells that had been treated with retinoic acid and dibutyryl cAMP for 5 days no longer expressed SSEA-1 and had a completely different morphology, the cells becoming flattened, stellate and separated, characteristic of parietal endoderm. The morphology and SSEA-1 expression status of all EC cell populations were tested in this manner prior to harvesting of the cells for extract preparation.

### 3.2 CONTROLS FOR INTEGRITY AND COMPARABILITY OF WHOLE-CELL EXTRACTS

F9 EC and parietal endoderm-like extracts for use in comparative experiments were prepared simultaneously using the same preparations of solutions to minimize variation in extraction efficiency. Whole-cell extracts should be capable of supporting transcription *in vitro*.  

107
Fig. 3.1. Anti-SSEA1 staining of F9 EC cells and their differentiated derivatives. Monolayers of F9 EC cells or cells that had been differentiated for 5 days as described in section 2.10. were stained with anti-SSEA-1 antibody as described in section 2.11.

3.1.A. F9 EC cells stained with the brown reaction product of the peroxidase catalyzed reaction indicating the presence of SSEA-1. Magnification is x100.

3.1.B. Parietal endoderm-like cells, the differentiated derivatives of F9 cells, not stained with the brown reaction product indicating the absence of SSEA-1. Approximately 5% of the cells remained undifferentiated. Magnification is x100. A control reaction using no anti-SSEA-1 antibody did not produce brown stained cells with either cell type.
Fig. 3.2. Whole-cell extracts can support transcription *in vitro*. A primer extension analysis of *in vitro* transcription reaction products is shown. The DNA template for the transcription reaction was 0.5μg of the HPRTcat plasmid (Table 2.5). Whole-cell extracts of F9 EC cells or parietal endoderm-like cells (PE) were titrated as indicated. Tracks 1 and 6 used 60μg of protein, tracks 2 and 5 used 40μg and tracks 3 and 4 used 20μg. The markers were pBR322 digested with *Hinf*I. The major extension product is indicated.
If any of the proteins involved in the transcription complex have been significantly degraded the extract will not be able to support efficient transcription. Extracts were therefore tested in transcription reactions to confirm their integrity.

The template DNA used for \textit{in vitro} transcriptions was pHPRTcat (see Table 2.5); the HPRT promoter was chosen because HPRT is a "housekeeping" gene, \textit{i.e.} it is ubiquitously expressed, and therefore the promoter should be active in extracts from both undifferentiated cells and their differentiated derivatives. Fig. 3.2 shows a primer extension analysis of mRNA transcribed \textit{in vitro} using pHPRTcat DNA as a template in either F9 EC or parietal endoderm-like cell extracts. Comparison with DNA marker fragments of known size showed the major extension product from both extracts to be 124 bp. This was the expected size, corresponding to a transcript from the known major transcriptional start of HPRT (Melton et al., 1986); thus both extracts were capable of accurate initiation of transcription. Approximately equivalent levels of transcription occurred with equal concentrations of each extract (part of the sample in track 4 was unfortunately lost during precipitation) suggesting that the extracts contained roughly equivalent concentrations of general transcription factors.

A second method of comparing the relative integrity of extracts was to use gel retardation to probe for a protein known not to be regulated during differentiation. One such factor is a protein which interacts with an Activating Transcription Factor (ATF) binding site (SivaRaman et al., 1986). Five ATF site binding activities have been defined in F9 cells, one of which, ECRE2, is not regulated during differentiation. (Tassios and La Thangue, in press).

Fig. 3.3.A shows a gel retardation experiment using the 'P' element oligonucleotide (for sequence see section 2.2), which includes a consensus ATF binding site, to probe an F9 EC cell extract. Two major retarded complexes formed in the presence of F9 EC cell extract (track 2). Competition analysis was used to investigate the specificity of each complex for the probe. In all competition experiments described the competitors were unlabelled DNA fragments or double-stranded oligonucleotides; competitor DNA was incubated
Fig. 3.3. F9 EC and parietal endoderm-like extracts contain equivalent concentrations of a non differentiation-regulated factor.

3.3.A. The 'P' element oligonucleotide was used to probe 3μg of F9 EC cell extract. 10ng of unlabelled competitor DNAs were added where indicated. In all reactions 0.5μg of sonicated salmon sperm DNA were used as a non-specific competitor. The complex involving ECRE2 (CRE2) is indicated.

3.3.B. The 'P' element oligonucleotide was used to probe 3μg of two F9 EC cell extracts, and two parietal endoderm-like extracts (F9PE; the differentiated derivatives of F9 cells). The complex involving ECRE2 (CRE2) is indicated.
with protein extract and sonicated salmon sperm DNA prior to addition of the probe. If unlabelled probe competed for complex formation more efficiently than an equivalent concentration of an unrelated sequence then the complex was concluded to be specific for that probe. Fig. 3.3.A shows that 10ng of unlabelled ‘P’ element oligonucleotide (track 3) competed more efficiently for the slower migrating complex than 10ng of an oligonucleotide of unrelated sequence (track 4). This slower migrating complex was therefore specific. Conversely, the two competitors competed equally efficiently for the faster migrating complex, which was therefore judged to be non-specific (compare tracks 3 and 4).

A non-specific complex, of equivalent mobility to the non-specific complex formed with the ‘P’ element, was formed with all probes and with extracts from all cell types tested. However, the addition of the non-specific competitor, sonicated salmon sperm DNA, to all gel retardation experiment reactions prevented the formation of large multi-component non-specific complexes. Presumably such complexes involve histone proteins, proteins involved in DNA replication, or other proteins able to interact with DNA in a non-sequence-specific fashion.

The specific complex identified in Fig. 3.3.A corresponded to the ‘P’ element interacting with ECRE2, the non-differentiation regulated ATF site binding activity. Fig. 3.3.B shows a gel retardation experiment in which the ‘P’ element was used to probe two F9 EC cell extracts and two parietal endoderm-like extracts. Each extract tested showed approximately equal signal intensities of the ECRE2 complex, once again confirming the comparability of these extracts.
3.3 GEL RETARDATION EXPERIMENTS USING RESTRICTION FRAGMENTS OF THE Mo-MSV LTR AS PROBES

The sequence of the Mo-MSV 5' LTR is shown in Fig. 3.4. Three sites of particular interest are indicated at -166, -327, and -345 bp upstream of the transcriptional start. These are the sites of point mutations in MPSV and PCMV, derivatives of Mo-MuLV able to be transcribed in EC cells.

-345 :- C -> T
-327 :- T -> A
-166 :- A -> C

Other mutations from the Mo-MSV sequence are not indicated because they are reversions to the Mo-MuLV sequence. Mo-MSV is derived from Mo-MuLV and neither can be transcribed in EC cells. MPSV has the point mutations at -166 and -327, PCMV has the same two point mutations plus a point mutation at -345 and a deletion of one of the 75 bp direct repeat sequences (see Chapter 1, section 1.5.4 for a more detailed description of the structures of these viruses).

Two restriction fragments were prepared from the Mo-MSV LTR, from DdeI (-372) to DdeI (-280), and from PvuII (-252) to Xbal (-152). Together these fragments encompass all three of the sites where the mutations have occurred, and the region deleted in PCMV. The proteins from EC cells and their differentiated derivatives that interact with these fragments were therefore compared using gel retardation analyses.

Fig. 3.5 shows a gel retardation experiment using the -252 to -152 sequence to probe extracts from F9 EC cells, their differentiated derivatives, PCC4 EC cells and HeLa cells (a human epithelial cell line). In each extract two major complexes of identical mobility were formed. The signal intensities for each complex were approximately equivalent in F9 EC and parietal endoderm-like extracts (compare tracks 2 and 3). This result shows that the complexes which form with the Mo-MSV sequence from -252 to -152 are not differentiation regulated.

Fig. 3.6.A shows a gel retardation experiment using the -372 to -280 fragment to probe extracts from a variety of different cell
Fig. 3.4. Sequence of the Mo-MSV 5’ LTR. Restriction sites used in the generation of probes for gel retardation experiments are indicated. The sites of mutations in the mutant viruses MPSV and/or PCMV are numbered (numbering is with respect to the Mo-MSV transcriptional start). The region encompassing one direct repeat sequence that is deleted in PCMV is indicated in lower case. The sequences of synthetic oligonucleotides used as probes in gel retardation experiments are underlined and numbered.

1 = Sp1 (-176 to -157)
2 = NF-1 (-333 to -309)
3 = MSV (-363 to -333)
types. With the F9 and PCC4 EC cell extracts two retarded complexes a and b were formed (tracks 3 and 5), but with the differentiated extracts only the slower migrating complex, a, formed (tracks 1, 2, and 4). Complex b has a lower signal intensity when formed with PCC4 cell extract than with F9 cell extract (compare tracks 3 and 5). As complex b was apparently EC cell specific a further gel retardation experiment was carried out using the -372 to -280 fragment to probe three different F9 EC cell extracts and two different parietal endoderm-like extracts. Fig. 3.6.B confirms that complex b forms only with extracts of EC cells (compare tracks 1-3 with tracks 4-5).

3.4 GEL RETARDATION EXPERIMENTS USING SYNTHETIC OLIGONUCLEOTIDES

Oligonucleotides equivalent to Mo-MSV sequences, or the corresponding sequences from MPSV or PCMV, around the sites of the mutations shown in Fig. 3.4 were synthesized. It was predicted that the wild-type oligonucleotide sequences might interact with an EC cell repressor of transcription; in such a scenario the mutant sequence would be predicted to have a lower affinity for the same factor. Conversely, it was predicted that the mutant sequences might interact with an activator (or activators) of transcription; in this case the wild-type sequence would be predicted to have lower affinity for the same factor(s). As with the restriction fragment probes, the proteins from EC cells and differentiated cells interacting with these oligonucleotides were investigated by gel retardation analyses.

The sequence around the -345 mutation, unique to PCMV, shows no obvious homology to a known transcription factor binding site. The oligonucleotides representing the Mo-MSV and PCMV sequences in this region have been termed MSV and PCMV respectively. The sequence around the -327 mutation has some homology with the consensus binding site of the transcription factor NF-1 (Nagata et al., 1982). The oligonucleotide representing the Mo-MSV sequence in this region has therefore been termed NF-1 WT, and the oligonucleotide representing the corresponding PCMV (or MPSV) sequence has been termed NF-1 mut.

The PCMV (or MPSV) sequence immediately surrounding the -166 mutation forms a consensus binding site for the well characterized
Fig. 3.5. Non-differentiation regulated complexes form with the Mo-MSV LTR between -252 and -152 bp with respect to the transcriptional start. The *Pvu*II (−252) to *Xba*I (−152) fragment of the Mo-MSV LTR was used to probe 3μg of various whole-cell extracts. F9 EC cells, their differentiated derivatives (parietal endoderm-like cells; F9PE), PCC4 EC cells and HeLa cells (a transformed human epithelial cell line). In each reaction 0.5μg of sonicated salmon sperm DNA were used as a non-specific competitor. The two major retarded complexes are indicated.
Fig. 3.6. An EC cell-specific complex forms on the Mo-MSV LTR between -372 and -280 bp with respect to the transcriptional start.

3.6.A. Gel retardation experiment using the $DdeI$ fragment from -372 to -280 bp of the Mo-MSV LTR to probe 3μg of various extracts. HeLa, SV3T3 clone 49 (a transformed murine fibroblast cell line; 3T3), PCC4 EC cells, parietal endoderm-like cells (PE), and F9 EC cells. In each reaction 0.5μg of sonicated salmon sperm DNA were used as a non-specific competitor. The two retarded complexes a and b are indicated.

3.6.B. Gel retardation using the $DdeI$ fragment to probe 3μg of three different F9 EC cell extracts and two different parietal endoderm-like (PE) differentiated F9 cell extracts. The two retarded complexes a and b are indicated.
However, the non-specific complex, a, is also competed more efficiently by wild-type MSV oligonucleotide than by the mutant PCMV oligonucleotide. From this single competition experiment it cannot be concluded that the wild-type sequence has a higher affinity for the EC cell-specific protein. On repeating the experiment it became apparent that the wild-type sequence does indeed have higher affinity for this protein (see Chapter 4, in particular Fig. 4.2).
mammalian activator of transcription Sp1 (Kadonaga et al., 1986). However, the corresponding Mo-MSV sequence has a single base-pair difference from the consensus. The oligonucleotide representing the PCMV (or MPSV) sequence around -166 has therefore been termed Sp1 WT, and the oligonucleotide representing the corresponding Mo-MSV sequence has been termed Sp1 mut.

The MSV oligonucleotide was used to probe extracts of EC cells and their differentiated derivatives. Fig. 3.7.A shows that one complex, b, only formed with EC cell extracts and the MSV probe. Two different F9 EC cell extracts and a PCC4 EC cell extract formed complex b (tracks 2, 3, and 6). The signal intensity of complex b is lower when formed with a PCC4 extract (track 6) than when formed with F9 cell extracts. Conversely, two different extracts of parietal endoderm-like cells (the differentiated derivatives of F9 cells) and a differentiated PCC4 cell extract did not form complex b (tracks 4, 5 and 7). Complex a, of slower mobility than b, formed with all the different extracts tested and had equivalent mobility to the non-specific complex formed with the 'P' element oligonucleotide (discussed above). Complex c, of slower mobility than a, only formed with the differentiated, parietal endoderm-like cell extracts (tracks 4 and 5).

The PCMV oligonucleotide, with a single point mutation at -345, was also used to probe an F9 EC cell extract. Fig. 3.7.B shows that the same two complexes, a and b formed on both the MSV and PCMV probes. However, the signal intensity for complex b was lower with the PCMV oligonucleotide; the signal intensity for complex a did not vary between the two different probes (compare tracks 2 and 10). A competition experiment is also shown in Fig. 3.7.B. Complex b was competed more efficiently when the wild-type MSV oligonucleotide was used as a competitor than when an equal concentration of the mutant PCMV oligonucleotide was used (compare tracks 3-5 with 6-8, and 11-13 with 14-16).

A competition experiment was also carried out using the MSV probe and a parietal endoderm-like cell extract, to investigate the specificity of complex c. Fig. 3.7.C shows that the Sp1 WT
Fig. 3.7. An EC cell specific complex forms with the Mo-MSV sequence from -363 to -333 with higher affinity than with the corresponding PCMV sequence.

3.7.A. The MSV oligonucleotide was used to probe 3μg of two F9 EC, two parietal endoderm-like (PE), PCC4, and differentiated PCC4 (PCC-D) cell extracts. In each reaction 0.5μg of sonicated salmon sperm DNA were used as a non-specific competitor. The three complexes formed, a, b and c, are indicated.

3.7.B. The MSV and PCMV oligonucleotides were used to probe 3μg of an F9 EC cell extract. Unlabelled MSV and PCMV oligonucleotides were used as competitors as indicated.
3.7.C. The MSV oligonucleotide was used to probe 3μg of a parietal endoderm-like cell extract. Unlabelled MSV, PCMV and Sp1 WT oligonucleotides were used as competitors as indicated.
oligonucleotide did not compete for complex c (tracks 9-11) showing it to be specific to the probe. However, MSV and PCMV oligonucleotides competed equally efficiently for complex c (compare tracks 3-5 with 6-8).

The NF-1 WT oligonucleotide was also used to probe extracts of EC cells and their differentiated derivatives. Fig 3.8 shows that multiple complexes formed with both F9 EC and parietal endoderm-like extracts. Some of the complexes are shared by both extracts and others are apparently specific to one extract. Competition with unlabelled NF-1 oligonucleotides and an unrelated sequence (Spl WT) showed that several of the complexes were specific to the probe (compare tracks 3 and 7, and tracks 10 and 14), but that none of the complexes were differentially competed by the wild-type or mutant NF-1 sites. NF-1 WT and mut oligonucleotides competed with equal efficiency for all the complexes (compare tracks 2 and 3 with 4 and 5, and tracks 9 and 10 with 11 and 12). This result indicates that the -327 point mutation does not significantly affect complex formation. This experiment was performed several times, with different probe preparations, but it did not prove possible to further separate the multiple complexes.

Fig. 3.9.A shows a gel retardation experiment in which the Spl WT oligonucleotide was used to probe a variety of extracts from differentiated and undifferentiated cell types. Four complexes A, B, C and D formed with extracts of F9 and PCC4 EC cells, their differentiated derivatives, and HeLa cells (the human cell line from which Spl has been purified and cloned). There were small variations in the relative signal intensities from the four complexes, but the electrophoretic mobilities were identical between the different cell extracts. Fig. 3.9.B shows a competition analysis of a gel retardation in which the Spl WT and Spl mut oligonucleotides were used to probe an F9 EC cell extract. The Spl WT probe (PCMV -176 to -157) interacted with the F9 EC cell extract to produce complexes A-D as before (track 2). Comparison of competition with the PCMV sequence (Spl WT) and an unrelated sequence (tracks 3 and 5) showed that complex D was not specific to the probe as it was competed to an equivalent degree by both sequences. Complexes A-C were
Fig. 3.8. Gel retardation analysis using the NF-1 WT oligonucleotide. NF-1 WT oligonucleotide was used to probe 3μg of F9 EC and parietal endoderm-like (F9 PE) cell extracts. 0.5μg of sonicated salmon sperm DNA were included in each reaction as a non-specific competitor. Unlabelled NF-1 WT, NF-1 mut and Spl WT oligonucleotides were used as competitors where indicated.
Fig. 3.9. The -176 to -157 sequence forms three specific complexes with a variety of extracts.

3.9.A. Gel retardation experiment using the PCMV sequence from -176 to -157 to probe extracts from a variety of cell types. F9 EC cells, parietal endoderm-like cells (F9PE; the differentiated derivatives of F9 cells), PCC4 EC cells, differentiated PCC4 cells (PCC-D) and HeLa cells. 0.5\(\mu\)g of sonicated salmon sperm DNA were included in each reaction as a non-specific competitor. Four retarded complexes A-D are indicated.

3.9.B. Gel retardation experiment using the PCMV sequence from -176 to -157 and the corresponding Mo-MSV sequence (MSV) labelled to an equivalent specific activity, to probe an F9 EC cell extract. Competitors were the unlabelled probes or the -363 to -333 Mo-MSV sequence as indicated.
specific to the probe as they were competed more efficiently by the Spl WT sequence. Comparison of competition with the PCMV (Spl WT) and Mo-MSV (Spl mut) sequences (tracks 3 and 4) suggests that complexes A-C formed with a higher affinity on the Spl WT probe. This result was confirmed by using the Spl mut oligonucleotide as a probe (tracks 5 to 10). Although complex D had an equivalent signal intensity with PCMV (Spl WT) and Mo-MSV (Spl mut) probes, complexes A-C had approximately 10-fold lower signal intensity when the Spl mut probe was used (compare tracks 2 and 7). Similarly, the PCMV sequence competed more efficiently than the Mo-MSV sequence for complexes A-C (compare tracks 7 and 8).

3.5 DISCUSSION

Whole-cell extracts were made from F9 EC cells, their differentiated derivatives and several other cell types by the method of Manley et al. (1980). The integrity of the extracts was confirmed by showing them to be capable of supporting in vitro transcription from the promoter of a "house-keeping" gene. Gel retardation experiments with an oligonucleotide probe representing an ATF binding site, showed that the extracts contained equivalent concentrations of a non-differentiation regulated DNA binding protein, ECRE2. These extracts were used to identify proteins from EC cells and differentiated cells which interact with the LTR sequences of Mo-MSV and its mutant derivatives.

It was predicted that EC cells might contain a repressor of transcription able to interact with wild-type Mo-MSV sequences, or that EC cells might lack activators of transcription that can interact with the Mo-MSV LTR. Gel retardation experiments using restriction fragments of the Mo-MSV LTR to probe a variety of cell extracts identified an EC cell-specific complex which forms with the wild-type LTR sequence between -372 and -280. This region includes the sites of the -345 point-mutation, which is unique to PCMV, and the -327 point mutation shared by MPSV and PCMV (this sequence occurs once in each direct repeat of the LTR). As this complex forms only with extracts from EC cells, in which Mo-MSV cannot be transcribed,
its protein component is a candidate for a repressor of transcription.

A fragment from -252 to -152 formed two major complexes, which had equivalent mobility in all extracts tested. Similar signal intensities were obtained for the two major complexes, and also for the minor complexes detected, using extracts of F9 EC cells and their differentiated derivatives. This restriction fragment includes the sites of the -166 point mutation shared by the MPSV and PCMV mutant retroviruses, and the sequences deleted in PCMV. As no complexes form with this region which are specific to either EC cells or to differentiated cells, it is unlikely that any EC cell-specific repressors of transcription or differentiated cell-specific activators of transcription interact with this region of the wild-type virus.

Similar experiments were performed using synthetic oligonucleotides to probe extracts from a variety of different cell types. The use of oligonucleotides enabled comparison of the proteins interacting with wild-type and mutant sequences. Three complexes, a, b and c, formed with the Mo-MSV sequence around -345 bp. Complex a is non-regulated during differentiation, and non-specific to the probe. Complex c is parietal endoderm-like cell-specific, and is specific to both the Mo-MSV and PCMV sequences. This complex could involve a differentiated cell-specific activator of transcription.

Complex b was EC cell-specific, similar to the complex b which formed on the -372 to -280 bp DNA fragment. The two b complexes also had similar electrophoretic mobility, and both had lower signal intensity when formed with PCC4 cell extract than with F9 cell extract. As the fragment includes the sequence of the oligonucleotide probe, and the complexes they form share common properties, it is likely that these complexes share a common protein component. When the corresponding PCMV sequence, with a single base-pair change at -345, was used to probe an extract of F9 EC cells, complex b formed but a lower signal intensity was detected. Consistent with this result, competition analyses showed that the wild-type sequence formed complex b with a higher affinity than the corresponding PCMV sequence. Thus the protein component of complex b is EC cell-specific.
and has a higher affinity for the wild-type, EC cell-repressed LTR sequence, i.e. the protein shows the DNA-binding properties predicted for a repressor of transcription.

The proteins interacting with the NF-1 WT and NF-1 mut oligonucleotides were compared by competition analyses in gel retardation experiments. These experiments showed that the point mutation at -327 had no affect, at either a qualitative or a quantitative level, on the DNA-protein complexes formed. However, a comparison of the complexes formed by the NF-1 oligonucleotides with extracts of F9 EC cells or their differentiated derivatives revealed some differences in the binding patterns. It did not prove possible to clearly resolve the multiple complexes which formed, and it was therefore difficult to further investigate the differences in the EC and differentiated cell proteins interacting with this site. These proteins may nevertheless play a role in the inability of Mo-MSV to be transcribed in EC cells.

The Spl WT oligonucleotide includes a consensus binding site for the mammalian activator of transcription Spl and corresponds to the PCMV and MPSV sequences around -166. This probe forms three specific complexes, A, B and C, with all extracts tested. It is possible that the protein component of one of these complexes involves a murine form of Spl protein. The corresponding Mo-MSV sequence forms the same complexes but with approximately 10-fold lower affinity. This result suggests that the protein components of these complexes are involved in activation of transcription, as they have a higher affinity for the mutant, EC cell expressed, LTR sequences.

In summary, two point mutations found in mutant viruses capable of being transcribed in EC cells, have proved to affect the interactions of cellular proteins with the retroviral LTR. The PCMV mutation at -345 bp (C -> T) reduces the binding affinity for an EC-cell specific complex. The mutation shared by PCMV and MPSV at -166 bp (A -> C) increases the binding affinity for three ubiquitous complexes. The further analysis of each of these mutations is considered in Chapters 4 and 5, respectively.
CHAPTER 4

ANALYSIS OF THE -345 MUTATION OCCURRING IN THE PCMV MUTANT RETROVIRUS
4.1 INTRODUCTION

Gel retardation experiments identified an EC cell complex, complex b, which formed with a restriction fragment of the Mo-MSV LTR from -372 to -280 bp with respect to the transcriptional start (Chapter 3, section 3.3). An EC cell specific complex with similar electrophoretic mobility also formed on a synthetic oligonucleotide corresponding to the Mo-MSV LTR sequence from -363 to -333 (oligonucleotide MSV; the Mo-MuLV sequence is identical in this region). This complex formed with a lower affinity on the equivalent PCMV sequence (oligonucleotide PCMV) which has a single base-pair difference at -345 bp. These results are shown in Chapter 3, section 3.4.

It has been postulated that the inactivity of the Mo-MSV LTR in EC cells is due to the presence of a repressor of transcription (Gorman et al., 1985a; 1985b). Such a repressor would have to be present in an active form only in EC cells. It might also be expected that such a repressor would be unable to interact with mutant derivatives of Mo-MSV which can be transcribed in EC cells. Complex b, described above, apparently fulfills both of these predictions and is therefore an excellent candidate for a repressor of transcription. In order to further investigate this possibility, the DNA binding properties of complex b were further investigated.

4.2 COMPLEX b INVOLVES THE SEQUENCE BETWEEN -372 AND -329

Fig. 3.6 showed that two complexes, a and b, formed in gel retardation experiments using the DdeI fragment from -372 to -280 of the Mo-MSV LTR and EC cell extracts, but only complex a formed when extracts from differentiated cell types were used. In order to address the specificity of these complexes a competition experiment was performed. Fig. 4.1 shows a gel retardation experiment using the 92 bp DdeI fragment as probe in an F9 EC cell extract. The competitor DNAs used were a 220 bp fragment extending from -539 of the Mo-MSV LTR to -327 and into the cloning vector of pMSVcat, and a 220 bp AvaI
Fig. 4.1. Competition analysis of complexes a and b formed on the Mo-MSV sequence from -372 to -280 using DNA fragments. The 92 bp DdeI fragment from -372 to -280 of the Mo-MSV enhancer was used to probe an F9 EC cell extract. 0.5μg of sonicated salmon sperm DNA were added to each reaction as a non-specific competitor. Specific competitors were the unlabelled 220 bp XbaI-PvuII fragment from -549 (within the cloning vector of pMSVcat) to -329 and an unlabelled 220 bp AvaI fragment from pUC19.
<table>
<thead>
<tr>
<th>Competitor (ng)</th>
<th>pUC220bp</th>
<th>MSV -549/-329</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extract</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

1 10 100

1 10 100

Extract: + + + + + + + +

Diagram with bands labeled a and b.
fragment from the cloning vector pUC19. 100ng of either competitor competed with equal efficiency for complex a, but 100ng of the Mo-MSV sequence competed several-fold more efficiently for complex b than 100ng of the pUC19 sequence (compare lanes 5 and 8). This result suggests that complex a is non-specific but that complex b is specific to the Mo-MSV sequence between -372 and -329 (the region over which competitor and probe overlap). Alternatively, binding sites for the same protein could occur in the non-overlapping regions of the probes.

4.3 THE PCMV SEQUENCE AROUND -345 FORMS COMPLEX b WITH APPROXIMATELY 10-FOLD LOWER AFFINITY THAN THE EQUIVALENT WILD-TYPE SEQUENCE

The DNA sequence involved in the formation of complex b was further defined by showing that the complex was also competed for by the MSV oligonucleotide corresponding to the Mo-MSV sequences from -363 to -333. Fig. 4.2 shows a gel retardation experiment once again using the DdeI fragment from the Mo-MSV LTR to probe an F9 EC cell extract. The competitors used were the -363 to -333 sequence from Mo-MSV and the equivalent PCMV sequence which has a single base-pair difference at -345. 1ng of the wild-type sequence competed efficiently for complex b but 10ng of the equivalent PCMV sequence was required to give an equivalent reduction in signal intensity. The two oligonucleotides competed for the non-specific complex a to an equivalent degree. From these results it may be concluded that complex b requires only those sequences between -363 and -333 of the Mo-MSV LTR for efficient formation, and that the same complex can also form on the equivalent PCMV sequence but with lower affinity.

4.4 OFF-RATE ANALYSIS CONFIRMS THAT COMPLEX b FORMS WITH HIGHER AFFINITY ON THE WILD-TYPE SEQUENCE AROUND -345 THAN ON THE CORRESPONDING PCMV SEQUENCE

In order to further quantitate the difference in the affinity of complex b when formed on the Mo-MSV or the PCMV sequence an "off-rate" analysis was performed. The method used is based on that of Hardy and Shenk (1989). The oligonucleotides MSV and PCMV, corresponding to the -363 to -333 sequences from the Mo-MSV and PCMV
Fig. 4.2. Competition analysis of complexes a and b formed on the Mo-MSV sequence from -372 to -280 using synthetic oligonucleotides. The 92 bp DdeI fragment from -372 to -280 of the Mo-MSV enhancer was used to probe an F9 EC cell extract. 0.5μg of sonicated salmon sperm DNA were added to each reaction as a non-specific competitor. Specific competitors were synthetic oligonucleotides corresponding to the PCMV sequence from -363 to -333 (PCMLTR) or the equivalent Mo-MSV sequence (MSVLTR).
Probe = Dde I-Dde I (-372/-280)

Extract = F9EC

Competitor (ng) =

<table>
<thead>
<tr>
<th></th>
<th>MSVLTR -363/-333</th>
<th>PCMVLTR -363/-333</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 10 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 10 100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Image of gel with bands labeled a and b]
Fig. 4.3. Off-rate analysis of complex b when formed on the Mo-MSV or PCMV sequences around -345.

4.3.A. 0.25ng of end-labelled MSV and PCMV oligonucleotides (MSVLTR and PCMVLTR), of equivalent specific-activity, were incubated with 3μg of F9 EC cell extract and 0.5μg of sonicated salmon sperm DNA for 15 min at 30°C as described for gel retardation experiments in section 2.20. The reaction mixtures were then chilled on ice and a 500-fold excess of unlabelled MSV oligonucleotide added. The reactions were transferred to 10°C and incubated for 1 to 20 min before returning to ice and analyzing by gel retardation. The reactions were started at timed intervals to allow simultaneous loading of the samples. A 6 h exposure of the MSV experiment and an 18 h exposure of the PCMV experiment are shown for ease of comparison.

4.3.B. Densitometric analysis of the off-rate assay. The intensity of the signal produced by complex b in each lane was measured by densitometry of autoradiographs using a Joyce-Loebl Chromoscan 3 densitometer. The values are expressed as a percentage of the signal intensity at time zero, following equalization with respect to the intensity of the signal produced by complex a. These values are shown plotted as a function of time.
A

Probe = MSVLTR-363/-333
Probe = PCMVLTR-363/-333

Competitor = 500 x Excess-363/-333
Competitor = 500 x Excess-363/-333

Time (minutes) 20 16 12 8 4 2 1 0

B

% COMPLEX b

0 1 2 3 4 8 12 16 20

TIME (mins)

○ Probe = MSVLTR-363/-333
● Probe = PCMVLTR-363/-333
LTRs, respectively, were end-labelled to equivalent specific-activity and used as probes in a gel retardation experiment. Each probe was incubated with 3μg of F9 EC cell extract to allow the formation of complex b and the reaction mixtures placed on ice. In order to determine the half-life of complex b on each sequence an excess of wild-type binding site, in the form of unlabelled MSV oligonucleotide, was added to the reaction mixtures. Presence of excess competitor ensures that once complex b, formed between the sequence-specific DNA binding protein and the labelled probe dissociates, it will not reform. Incubations in the presence of excess competitor were performed at 10°C, a temperature which was empirically determined to be most suitable as it slowed the reaction enough to allow monitoring of complex dissociation by gel retardation.

Fig. 4.3.A shows an example of an "off-rate" experiment. The dissociation rate of complex b was slower when the wild-type Mo-MSV sequence was used as the probe than with the PCMV sequence. This result confirms that shown in Fig. 4.2 which suggested that complex b had a higher affinity when formed on the wild-type sequence. To take into account the lower signal intensity obtained when complex b was formed with the mutant probe, an 18 h exposure of the gel using the PCMV probe is shown and a 6 h exposure of the gel using the wild-type probe. The two probes gave equivalent signal intensities for the non-specific complex a, once again confirming the lower affinity of complex b when formed on the mutant probe.

The non-specific complex a provides an internal control within each track of the experiment for loading and pipetting errors. Fig. 4.3.B is a graphical representation of the results in Fig. 4.3.A. The signal intensity of complex b was measured in each track by densitometry, and normalized by comparison to the signal intensity of complex a in that track. The results are expressed as a percentage of the signal intensity of complex b at time zero and plotted with respect to time. A line of best fit is shown for the values obtained with each probe. This experiment was repeated several times and a similar trend was always observed. The dotted lines indicate the time at which 50% of complex b remained on each probe, showing that the half-life of complex b is approximately 6-fold longer when the
complex is formed on the wild-type sequence than when it is formed on
the mutant sequence.

Taking into account the investigations described above, the protein
component of complex b continued to fulfill the requirements for an
EC cell specific repressor of transcription from the retroviral LTR.
The protein (or protein complex) is therefore termed EC-R for EC
cell-Repessor of transcription from this point onwards.

4.5 A SEARCH FOR SEQUENCES HOMOLOGOUS TO THE EC-R BINDING SITE

Any host cell protein interacting with a virus presumably has a
functional role during the normal life of the cell. Therefore it is
to be expected that EC-R binding sites will occur in the promoters of
cellular genes. However, it is difficult to make any extensive search
for such sites; firstly because the precise limits of the EC-R
binding site within the Mo-MSV LTR remain undefined, and secondly
because varying sequences may be equally capable of forming a
functional interaction with EC-R. A number of attempts to better
define the sequences bound by EC-R within the Mo-MSV LTR by DNAseI
footprinting were unsuccessful. No obviously protected sequence was
identified. This problem may reflect a low abundance of the EC-R
protein in the whole-cell extracts used, or a relatively low affinity
of interaction. During the course of this work Tsukiyama et a/. (1989)
reported the presence of an EC cell-specific protein binding
to the Mo-MuLV LTR in the -345 bp region. This protein, termed ELP
for embryonic LTR-binding protein, is presumably identical to the EC-
R protein. With nuclear extracts, rather than whole-cell extracts,
this group successfully defined the binding site for ELP using DNAseI
footprinting. The protected residues corresponded to the Mo-MSV
sequences from -345 to -351.

Taking into consideration the results of Tsukiyama et a/. (1989),
the EMBL sequence data base was searched for the 10 base sequence
motif CAAGGTCAGG, corresponding to the Mo-MSV sequences from -351
to -342, using the program gcgFIND and allowing no mismatches.
Despite the limitations of such an approach the sequence motif was
detected at 141 sites in the database, a significantly greater number
than that predicted for a random decamer sequence (the search encompassed 29 x 10^6 bases, a random decamer should occur 28 times in that length of sequence). Amongst the 141 sites were several retroviral LTRs including those of the Abelson murine leukaemia virus, the Friend murine leukaemia virus, the Rauscher mink cell focus forming virus and derivatives of these viruses. The motif also occurs within human Alu repeat sequences and within the coding sequence of multiple cellular and viral genes. More interestingly, the motif was also found in the promoters of the following genes: the human transferrin receptor, at approximately -300 bp from the transcriptional start; the murine nerve growth factor \( \alpha \) subunit, in opposite orientation at approximately -285 bp (also in the \( \gamma \) subunit); and the murine immunoglobulin V-\( \kappa \)10-Ars-A \( \kappa \) chain. However, the significance of the presence of the decamer motif in these promoters is not known.

The EC-R binding site was also compared by eye with the binding sites of known transcription factors. Partial homology was detected with the polyomavirus enhancer activator-2 (PEA-2; Piette and Yaniv, 1987) binding site, and with the polyomavirus enhancer-binding protein 5 (PEBP5; Asano et al., 1990) binding site.

\[-345\]

Mo-MSV : 5’ CAAGGTCAGGA 3’

PCMV : 5’ CAAGGT\( \text{t} \)AGGA 3’

PEA-2 site : 5’ TGGGTCAGTT 3’

*****

Mo-MSV : 5’ AAGGTCAGGAACAGAG 3’

PCMV : 5’ AAGGT\( \text{t} \)AGGAACAGAG 3’

PEBP5 site : 5’ ATAAACAGGAAGT\( \text{g} \)AC 3’

*****

The PEA-2 and PEBP5 binding sites as defined by footprinting (Piette and Yaniv, 1987; Asano et al., 1990) are underlined. Bases common to the Mo-MSV and Polyomavirus sequences are indicated by asterisks. The point mutation at -345 in the PCMV sequence is indicated by italics. Note that the PEA-2 site is shown in reverse orientation with respect to the PEBP5 site.
Fig. 4.4. Competition of the EC-R binding site with the polyomavirus A element. The 92 bp DdeI fragment from -372 to -280 of the Mo-MSV enhancer was used to probe an F9 EC cell extract in a gel retardation assay. 0.5μg of sonicated salmon sperm DNA were included in each reaction as a non-specific competitor. Specific competitors were synthetic oligonucleotides corresponding to the Mo-MSV and PCMV enhancer sequences from -363 to -333, the Py WT and Py mut sequences and the ATF binding site (for definitions and sequences see section 2.2.).
4.6 EC-R DOES NOT RECOGNIZE PEA-2 OR PEBP5 BINDING SITES

The EC-R binding site has 6 bp in common with the PEBP5 binding site. However, as PEBP5 concentrations have been reported to increase rather than decrease following differentiation of F9 EC cells (Asano et al., 1990), it was considered unlikely that PEBP5 is EC-R. PEA-2 has been implicated as a possible repressor of transcription (Wasylyk et al., 1988; see section 1.4.4). As the PEA-2 and EC-R binding sites have 6 bp in common and are both candidate transcriptional repressors it was hypothesized that PEA-2 is EC-R. In order to test this hypothesis a gel retardation experiment was performed in which the EC-R complex was competed by oligonucleotides corresponding to various factor binding sites. Fig. 4.4 shows the results of this experiment. The EC-R complex, b, is efficiently competed for by 0.5ng of unlabelled Mo-MSV -363 to -333 oligonucleotide. 0.5ng of the corresponding PCMV sequence do not compete as efficiently, as expected from earlier experiments. However, 5ng of the PCMV sequence completely compete out the EC-R complex.

The Py WT oligonucleotide includes the consensus binding sites for both PEA-2 and PEBP5. The Py mut oligonucleotide is the same sequence but with a single point mutation within the PEA-2 binding site at the equivalent base to the PCMV -345 mutation. The ATF oligonucleotide corresponds to an unrelated transcription factor binding site. The Py WT, Py mut and ATF oligonucleotides all compete for EC-R formation to an equivalent, very low level (Fig. 4.4, tracks 7 to 12). 5ng of the EC-R binding site oligonucleotide completely abolishes formation of complex b, but 5ng of the polyomavirus competitor has very little effect on the signal intensity of complex b. This result suggests that the PEA-2 site is unable to form a complex with the EC-R protein i.e. EC-R is not PEA-2.

4.7 HEPARIN-SEPHAROSE COLUMN CHROMATOGRAPHY AFFORDS 20-FOLD PURIFICATION OF EC-R

A partial purification of EC-R protein using heparin-Sepharose ion-exchange column chromatography was performed to facilitate further
Fig. 4.5. Partial purification of EC-R by heparin-Sepharose column chromatography. 1g of heparin-Sepharose was equilibrated in heparin-Sepharose column buffer (section 2.22) containing 0.1 M KCl. A Bio-Rad Econocolumn was then packed with the resin to give a bed volume of approximately 4ml. 15mg of F9 EC whole-cell extract were applied to the column which was washed with 5 column volumes of 0.1 M KCl in column buffer prior to application of a KCl gradient of 5 mM/min KCl until a final concentration of 1 M KCl was reached. The column flow-rate was 10 ml/h and 1.3ml fractions were collected.

4.5.A. Elution profile of the column. 10μl of each fraction were used in the Bio-Rad protein concentration assay.

4.5.B. Gel retardation analysis of column fractions using the MSV oligonucleotide corresponding to -363 to -333 of the Mo-MSV enhancer sequence. Approximately 3μg of protein from each fraction were used in the assay and 0.5μg of salmon sperm DNA were used as a non-specific competitor.
characterization of the protein. Fig. 4.5A shows the elution profile of F9 EC whole-cell extract from such a column. At low salt concentration greater than 60% of the protein remained bound to the column. On application of a salt gradient the majority of bound protein was eluted between 0.3 M and 0.5 M KCl concentration. 3μg of alternate fractions were analyzed for the presence of EC-R protein by gel retardation using the Mo-MSV -363 to -333 sequence as a probe. Fraction 23, which eluted at approximately 0.6 M KCl, contained the highest concentration of EC-R. The non-specific complex a, also elutes at this salt concentration. Comparison of EC-R signal intensity obtained with equivalent concentrations of unpurified and purified extract showed a 20-fold purification of EC-R in fraction 23. Competition analysis showed that the partially purified EC-R protein retained the same binding properties as EC-R from whole-cell extracts. Activity is maintained after ion-exchange chromatography, suggesting that EC-R consists of a single protein. However, multiple proteins components could stably interact during chromatography.

4.8 WHEAT GERM AGGLUTININ COLUMN CHROMATOGRAPHY SHOWS THAT EC-R IS NOT GLYCOSYLATED

The transcription factor Spl is heavily glycosylated with O-linked N-acetyl glucosamine residues (GlcNAc). It has been suggested that glycosylation may be a common phenomenon amongst transcription factor proteins (Jackson and Tjian, 1988). GlcNAc residues selectively bind to the lectin wheat germ agglutinin (WGA), which can be coupled to agarose and used to form chromatographic columns. The heparin-Sepharose fractions containing partially purified EC-R protein were applied to a WGA column and eluted with an excess of GlcNAc.

The elution profile of the column (Fig. 4.6.A) shows that the majority of protein applied was not bound. However, a small proportion of the protein was eluted on application of GlcNAc, showing that the heparin-Sepharose fractions contained some glycosylated protein. Fig. 4.6.B shows a gel retardation analysis of the fractions using the Mo-MSV -363 to -333 sequence as a probe. EC-R protein was detected in whole-cell extract, the heparin-Sepharose
Fig. 4.6. Wheat germ agglutinin column chromatography shows that EC-R is not glycosylated. 300μg of partially purified EC-R (heparin-Sepharose fraction; dialysed to 0.1 M KCl against WGA column buffer, section 2.23), were applied to 1ml of WGA coupled to agarose equilibrated in WGA column buffer. After washing with 5 column volumes of WGA column buffer, 1 M GlcNAc in column buffer was added. The column flow rate was 10 ml/h, 500μl fractions were collected.

4.6.A. Elution profile of the column. 50μl of each fraction were used in the Bio-Rad protein concentration assay. The point at which GlcNAc was added is indicated.

4.6.B. Gel retardation analysis of WGA column fractions. Fractions were dialysed to 0.1 M KCl against 100 volumes of Manley dialysis buffer (section 2.16). The MSV oligonucleotide corresponding to -363 to -333 of the Mo-MSV enhancer was used to probe 18μl of each fraction in the presence of 0.5μg of sonicated salmon sperm DNA as a non-specific competitor.
partially purified preparation, and in the WGA column flow-through fractions 2 to 6. However, no further EC-R protein was eluted by application of excess GlcNAc. This result suggests that the EC-R protein is not glycosylated with GlcNAc residues, or that an undetectable level of glycosylation exists. The WGA column was later shown to be capable of binding glycosylated proteins (see Chapter 5).

4.9 FPLC ANALYSIS SHOWS THAT EC-R HAS A NATIVE MOLECULAR WEIGHT OF ABOUT 60kd

The partially purified heparin-Sepharose fraction of EC-R protein was also used to estimate the native molecular weight of EC-R using fast protein liquid chromatography (FPLC) analysis, in conjunction with Dr. R. Hayes, Imperial College, London. Fig. 4.7.A shows a standard curve of molecular weight as a function of elution point from a Superose-12 FPLC column (Pharmacia). Heparin-Sepharose partially purified EC-R was then applied to the column; the elution profile is shown in Fig. 4.7.B. Some protein was detected in the void volume, but greater than 50% of the protein was detected in the elution volume, the majority eluting between fractions 10 and 25. Comparison with the standard curve in Fig. 4.7.A shows that this protein lies within the molecular weight range 40-400kd.

Alternate FPLC fractions were assayed for EC-R by gel retardation using the Mo-MSV -363 to -333 sequence. Fig. 4.7.C shows that EC-R activity peaked between fraction 22 and 24, corresponding to a molecular weight of approximately 60kd. Once again the retention of activity following chromatography suggests that EC-R is a single protein or a complex which remains stable in solution.

4.10 A SINGLE CYCLE OF AFFINITY CHROMATOGRAPHY AFFORDS A FURTHER 10-FOLD PURIFICATION OF EC-R

An affinity column was prepared by the method of Kadonaga and Tjian (1986), using concatamers of the MSV oligonucleotide representing the Mo-MSV sequence from -363 to -333. The efficiency of the ligation of
Fig. 4.7. Estimation of the native molecular weight of EC-R using FPLC.

4.7.A. A standard curve of molecular weight plotted as a function of elution point for a 1cm x 30cm Superose-12 FPLC column. Molecular weight standards were the Sigma MW-GF-200 kit. Approximately 0.2mg of each protein were applied to the Superose-12 column at a flow rate of 0.4 ml/min in heparin-Sepharose column buffer (section 2.22). The elution volume as a proportion of the void volume (Ve/Vo) is plotted with respect to the logarithm of the molecular weight.

4.7.B. Elution profile of partially purified EC-R (heparin-Sepharose fractions) from the Superose-12 column. 70μg of partially purified extract dialysed to 0.1 M KCl in heparin-Sepharose column buffer (section 2.22) were applied to the column at a flow rate of 0.4 ml/min and 200μl fractions collected. The eluate was passed through an U.V. monitor and protein concentration was determined by optical density measurement at 280nm, a full scale deflection corresponded to 0.2 A_{280} Units.
A

STANDARD CURVE OF MOLECULAR WEIGHT
PLOTTED AS A FUNCTION OF ELUTION POINT

\[
\frac{V_e}{V_0} \quad \text{cytochrome c.}
\]

\[
\text{carbonic anhydrase}
\]

\[
\text{albumin}
\]

\[
\text{alcohol dehydrogenase}
\]

\[
\text{beta-amylase}
\]

\[
dextran blue
\]

Log Molecular weight (dal.)

B

ELUTION PROFILE

\[
A_{280}
\]

Fraction
4.7.C. Gel retardation analysis of FPLC fractions. The MSV oligonucleotide corresponding to -363 to -333 from the Mo-MSV enhancer was used to probe 18μl of alternate FPLC fractions in the presence of 0.5μg sonicated salmon sperm DNA as a non-specific competitor.
Fig. 4.8. Further purification of EC-R by affinity chromatography. 4mg of partially purified EC-R (heparin-Sepharose fraction dialysed to 0.1M KCl in column buffer) were applied to an affinity column formed using concatemerized MSV (-363 to -333) oligonucleotide as described in section 2.24. The column volume was 5ml, the flow-rate 10 ml/h, 1.3ml fractions were collected. After extensive washing with 0.1 M KCl in column buffer a gradient of 5 mM/min KCl was applied to a final concentration of 1 M KCl.

4.8.A. Elution profile of the column. 50µl of each fraction were used in the Bio-Rad protein concentration assay.

4.8.B. Gel retardation analysis of column fractions. The MSV oligonucleotide corresponding to -363 to -333 of the Mo-MSV enhancer was used as the probe. Fractions were dialysed to 0.1 M KCl against 100 volumes of Manley dialysis buffer (section 2.16) and 18µl of fraction were used in each assay in the presence of 0.5µg sonicated salmon sperm DNA as a non-specific competitor.
oligonucleotides to form concatamers was assessed by agarose gel electrophoresis. The ligation reaction was successful producing up to 40-mers. The concatamers were then coupled to cyanogen-bromide activated Sepharose. 10ng of $^{32}$P-labelled concatamerized oligonucleotide were added to the reaction, thus allowing the course of the reaction to be followed. Greater than 50% of the oligonucleotide became coupled to the Sepharose.

Partially purified Sepharose fractions containing EC-R were applied to the column. Fig. 4.8.A shows the elution profile of the column. Approximately 20% of the protein bound to the column in 0.1 M KCl and was eluted on application of the salt gradient. Fig. 4.8.B shows a gel retardation analysis of the column fractions using the Mo-MSV -363 to -333 sequence to detect EC-R. No EC-R was detected in the flow-through, but fractions 18 to 20 contained EC-R activity showing that the protein had bound to the affinity column as expected. 3μg of whole-cell extract and 0.3μg of fraction 18 produced approximately equivalent EC-R signal intensities, showing that the affinity chromatography had afforded a further 10-fold purification of EC-R. The protein component of the non-specific complex co-purifies with EC-R throughout, although the majority elutes at a lower salt concentration than EC-R from the affinity column.

4.11 DISCUSSION

The results discussed above show that two different complexes form with the -363 to -333 bp region of the Mo-MSV LTR. Complex a forms with extracts of all cell types tested and is non-specific to the probe i.e. it can form with any DNA sequence. Conversely, complex b forms only with extracts of EC cells and is specific to the -363 to -333 sequence of Mo-MSV. Complex b can also form with the corresponding PCMV sequence but with at least 6-fold lower affinity. The PCMV sequence differs from the Mo-MSV sequence by just a single base-pair change at -345, and unlike the wild-type virus it is capable of being transcribed in both F9 and PCC4 EC cells. The protein component of complex b is therefore an excellent candidate for a repressor of transcription and has been termed EC-R for EC
cell-Repressor of transcription.

The competition experiments shown in Figs. 4.1 and 4.2 indicate that the EC-R complex formed between a 92 bp fragment and F9 EC cell extract was efficiently competed for by 1ng of -363 to -333 oligonucleotide. However, 100ng of the unlabelled fragment were required to efficiently compete for this complex. The effective concentration of binding sites is 3-fold higher for the oligonucleotide as the sequence is 1/3 of the length, leaving a 30-fold discrepancy.

An explanation is that reduction of the size of the binding site increases the affinity for EC-R. This could occur if a second protein binding site existed close to the EC-R site; by interacting with such a site the second protein could prevent EC-R binding e.g. by steric hindrance. However, no such interaction is detected by the gel retardation assay, and hence if such a complex forms it must have a very short half-life.

Some sequence homology exists between the EC-R binding site and the polyomavirus PEA-2 binding site, but competition experiments have shown that the EC-R protein is not equivalent to PEA-2. Sites with sequence homology to the EC-R binding site also occur in the LTRs of several other retroviruses. The reverse transcriptase enzyme has no proof-reading activity and the formation of DNA from the retroviral RNA genome is therefore error prone. Conservation of the EC-R binding site sequence between different retroviruses suggests that it plays a role in the retroviral life-cycle. However, this role may be secondary to any function as a transcription factor binding site.

Sites with homology to the EC-R binding site also occur in the promoters of several cellular genes. In each case the site occurs in a similar position relative to the transcriptional start to that found in the Mo-MSV LTR. This again may reflect a conservation of function. In order to determine whether these sites are functional, the effect of mutating the EC-R site on the level of transcription from these genes would have to be determined.
Using small-scale chromatography a purification protocol for the EC-R protein has been developed. Heparin-Sepharose ion-exchange chromatography affords a 20-fold purification, and affinity column chromatography a further 10-fold purification. One method of removing non-specific DNA binding proteins from the sample, including the protein component of complex a, would be to pass the partially purified extract down an affinity column prepared with the PCMV sequence from -363 to -333. This should bind the EC-R protein with a lower affinity than the column prepared with the equivalent Mo-MSV sequence, and hence EC-R would elute at a lower salt concentration leaving contaminating proteins on the column. Repeated rounds of chromatography using the affinity column prepared with the Mo-MSV sequence should then allow complete purification of EC-R to be achieved. By scaling up such a procedure enough protein could be obtained to enable partial sequencing of the EC-R protein and hence cloning of the EC-R gene.

The partially purified extract produced by heparin-Sepharose column chromatography was used in further characterization of the protein. Wheat germ agglutinin column chromatography showed that the EC-R protein is not glycosylated with N-acetyl glucosamine to a level detectable by this method. Although it has been suggested that glycosylation is a common property amongst transcription factors (Jackson and Tjian, 1988), only a few different factors have so far been tested for their glycosylation level. Lack of glycosylation does not, therefore, suggest that EC-R is not a transcription factor.

The partially purified extract was also used in FPLC analysis to show that the native molecular weight of the EC-R protein is around 60kd. This relatively low molecular weight, and the retention of binding activity during the various purification procedures, suggests that EC-R is a single protein. However, it is possible that it consists of two or more smaller proteins, which form a stable complex in solution in the absence of DNA. It is also possible that multimers of a single protein interact with DNA, to form the EC-R complex identified by gel retardation analysis. It is not possible to determine the size of a protein-DNA complex by gel retardation analysis, as migration rate is affected not only by molecular weight.
but also by charge and conformation of the complex.

In summary, EC-R is a previously uncharacterized, EC-cell specific, non-glycosylated, DNA-binding protein of approximately 60kd. It interacts with the LTR of Mo-MSV and Mo-MuLV, which are not transcribed in EC cells, but has at least 6-fold lower affinity for the corresponding sequence in PCMV, which is transcribed in EC cells. EC-R is therefore a strong candidate for an EC cell repressor of transcription.
CHAPTER 5

ANALYSIS OF THE POINT MUTATION AT -166 OCCURING IN MUTANT RETROVIRUSES CAPABLE OF BEING TRANSCRIBED IN EMBRYONAL CARCINOMA STEM CELLS.
5.1 INTRODUCTION

The PCMV and MPSV viruses share a point mutation at -166 (with respect to the transcriptional start) which generates a consensus binding site for the mammalian transcription factor Spl. Spl is a well characterized activator of transcription as discussed in section 1.3.3. Synthetic oligonucleotides, corresponding to the Mo-MSV and PCMV sequences from -176 to -157, were used in gel retardation experiments. The sequences of these oligonucleotides are shown in section 2.2. The PCMV sequence, which includes a consensus Spl binding site, is termed Spl WT and the corresponding Mo-MSV sequence, which differs from a consensus Spl binding site by a single base change at -166, is termed Spl mut.

In Chapter 3 the complexes formed by different whole cell extracts on the PCMV and Mo-MSV oligonucleotides were compared. Fig. 3.9.A showed that four complexes A-D formed on the PCMV (Spl WT) oligonucleotide with all the whole cell extracts tested. Fig. 3.9.B showed that complexes A-C were specific to the probe but complex D was non-specific. When the Mo-MSV (Spl mut) oligonucleotide was used as a probe with F9 EC cell extract the same pattern of complexes was produced. However, the concentrations of complexes A, B and C were reduced by approximately 10-fold (see section 3.5 for further discussion of these results).

The preliminary experiments described in section 3.5 suggest that the PCMV (and MPSV) LTR sequence, around -166, forms complexes with F9 EC cell proteins with a higher affinity than the corresponding Mo-MSV (and Mo-MuLV) LTR sequence. One or more of these complexes may be involved in activation of transcription. Hence generation of the Spl binding site at -166 could be involved in relieving the EC cell block to Mo-MSV transcription.

In order to further investigate this possibility it was important to further quantitate the differences in the ability of the PCMV and Mo-MSV sequences to form complexes A-C with F9 EC cell extract. The complexes themselves were also further characterized and because they form at an Spl binding site their protein components were compared to Spl protein.
5.2 THE PCMV SEQUENCE AROUND -166 COMPETES FOR COMPLEXES A TO C MORE EFFICIENTLY THAN THE CORRESPONDING Mo-MSV SEQUENCE

A more extensive competition analysis, involving a titration of competitor oligonucleotides, was carried out in order to further quantitate the differences in the affinities of complexes A-C when formed on the Mo-MSV or PCMV LTRs. The PCMV (Sp1 WT) oligonucleotide was used as a probe in an F9 EC cell extract and 1ng, 10ng or 100ng of competitor oligonucleotides were added to the reactions. At equal concentrations of competitor DNA the PCMV sequence competed for complexes A-C more efficiently than the Mo-MSV sequence (Fig. 5.1, compare lanes 4 and 7, 5 and 8, 6 and 9). A nonsense oligonucleotide did not compete for these complexes over this range of competitor concentration (Fig. 5.1, lanes 1-3). Using this assay 1ng of PCMV (Sp1 WT) oligonucleotide competed for complexes A-C as efficiently as 10ng of Mo-MSV (Sp1 mut) oligonucleotide.

5.3 THE PCMV SEQUENCE FROM -176 TO -157 HAS APPROXIMATELY 5-FOLD HIGHER AFFINITY FOR COMPLEXES A, B AND C THAN THE EQUIVALENT Mo-MSV SEQUENCE

A gel retardation experiment, in which the PCMV sequence was used as a probe and unlabelled PCMV and Mo-MSV oligonucleotide competitors were titrated, was used to more accurately quantitate the differences in binding affinity for the two probes (Fig. 5.2.A). At each competitor concentration complexes A-C were more reduced in response to competition with the PCMV sequence than with the Mo-MSV sequence. Densitometric analysis showed that approximately 5-fold more Mo-MSV competitor was required than PCMV competitor to produce an equivalent reduction in signal intensity (Fig. 5.2.B).

The results of these competition experiments show that complexes A-C form at the consensus Sp1 site in the PCMV LTR and at the corresponding site in the Mo-MSV LTR. The single base change from the consensus Sp1 binding site, found in the wild-type sequence, reduces the affinity of these interactions by at least 5-fold.
Fig. 5.1. Competition analysis of the complexes formed on the PCMV (Spl WT) probe. The PCMV oligonucleotide was used to probe 3µg of an F9 EC cell extract. 0.5µg of sonicated salmon sperm DNA were included in each reaction as a non-specific competitor. The specific competitors were unlabelled probe (PCMV), the corresponding Mo-MSV sequence (MSV) or an unrelated ("nonsense") oligonucleotide.
<table>
<thead>
<tr>
<th>Probe</th>
<th>PCMV -176/-157</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extract</td>
<td>F9EC</td>
</tr>
<tr>
<td>Competitor (ng)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nonsense</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

![Image of gel electrophoresis result]
Fig. 5.2. Competition analysis of the complexes formed on the PCMV probe by a titration of unlabelled Mo-MSV and PCMV probes.

5.2.A. Gel retardation assay using 3µg of F9 EC cell extract, 0.5µg of sonicated salmon sperm DNA, and an increasing concentration of each competitor. P = PCMV oligonucleotide, M = Mo-MSV oligonucleotide.

5.2.B. Densitometric analysis. The intensities of the three specific complexes A-C in each lane were measured by densitometry of the autoradiograph using a Joyce-Loebl Chromoscan 3 densitometer. Relative signal intensity, as a percentage of the intensity of the uncompeted signal, is plotted as a function of competitor concentration.
5.4 Complexes A-C are competed for by an HSV tk gene promoter fragment that includes two Sp1 binding sites

Complexes A-C formed preferentially on the PCMV sequence which includes a consensus Sp1 binding site. The protein component of one or more of these complexes may correspond to, or be related to, the Sp1 protein. In order to investigate this possibility competition experiments were performed using a DNA fragment containing previously characterized Sp1 binding sites. The Herpes simplex virus (HSV) thymidine kinase (tk) gene promoter, which includes two binding sites for transcription factor Sp1 (Kadonaga et al., 1986), was chosen for this purpose.

Fig. 5.3 shows a gel retardation experiment in which the PCMV (Sp1 WT) oligonucleotide was used as a probe and various concentrations of a fragment of the tk gene promoter, encompassing the two known Sp1 binding sites, was used as a competitor. An equivalent sized fragment of the pBR322 plasmid was used as a control competitor. 100ng of the tk fragment efficiently competed for complexes A-C, but 100ng of pBR322 fragment did not compete to any significant degree (Fig. 5.3, compare lanes 1 and 4). This result shows that the protein components of complexes A-C can also interact with the tk promoter, once again suggesting that they are related to Sp1 protein.

5.5 Complex A is recognized by anti-Sp1 antibodies

To further investigate which, if any, of complexes A-C is equivalent to Sp1, a polyclonal anti-Sp1 antibody, raised against human Sp1 protein, was added to gel retardation experiments. The anti-Sp1 antibody used was 2892-E (a kind gift from S. P. Jackson, Berkeley, USA). 1 or 2μl of antiserum or preimmune serum from the same rabbit were added to the binding reactions either before or after addition of the probe. A component of preimmune serum and the anti-Sp1 antibody preparation formed a novel retarded complex with the PCMV oligonucleotide probe, of slower mobility than complexes A-C (Fig. 5.4, lanes 3 to 12). When antibody and F9 EC cell extract were incubated together prior to addition of the probe, A', a distinct
Fig. 5.3. Complexes A-C are competed for by an HSV tk gene promoter fragment containing two Spl binding sites. The HSV tk gene promoter fragment is a 164 bp fragment from the BamHI to the BglII site of pBLCAT2 (Luckow and Shutz, 1987). The pBR322 fragment is a 234 bp fragment from BglII to BglII.
<table>
<thead>
<tr>
<th>Competitor (ng)</th>
<th>pBR322 Fragment</th>
<th>HSV tK Promoter</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Diagram showing Gel Electrophoresis with bands labeled A, B, C, and D.
Fig. 5.4. The protein involved in complex A is recognized by anti-Spl antibody. The PCMV (Spl WT) probe and an F9 EC cell extract were used in this assay. (-) and (+) refer to absence or presence of the extract. 1 or 2µl (as indicated) of anti-Spl antibody (Spl) or preimmune serum (P.I) were added prior to addition of the probe in tracks 2 to 6 or following addition of the probe in tracks 10 to 13.
novel complex was formed (Fig. 5.4, lanes 3 and 4). The concentration of complex \(A'\) increased as the concentration of complex \(A\) decreased on addition of antibody. Preimmune serum did not form complex \(A'\) (Fig. 5.4, lanes 5 and 6). When antibody or preimmune serum were added to the reaction after addition of the probe the same results were obtained (Fig. 5.4, lanes 10 to 13). These results show that the anti-Spl antibody interacts with the protein component of complex \(A\) to form a ternary complex \(A'\). Preincubation with anti-Spl antibody did not prevent formation of complex \(A\) suggesting that antibody binding and DNA binding require different regions of the protein. These results suggest that the protein component of complex \(A\) is immunologically related to human Spl protein.

5.6 PURIFIED Spl PROTEIN FROM HELEA CELLS FORMS A COMPLEX WITH THE PCMV OLIGONUCLEOTIDE OF EQUIVALENT MOBILITY AND BEHAVIOUR IN COMPETITION EXPERIMENTS TO COMPLEX A

Purified human Spl protein from HeLa cells (a kind gift from S. P. Jackson) was used in a gel retardation experiment with the PCMV (Spl WT) probe. 2.3ng of purified protein were used in each reaction and no non-specific competitor was used. The purified Spl formed a complex of equivalent mobility to complex \(A\) (Fig. 5.5, compare lanes 1 and 2). Competition experiments were also performed using the purified Spl protein. The PCMV (Spl WT) oligonucleotide competed for the single complex more efficiently than an equivalent concentration of the Mo-MSV (Spl mut) oligonucleotide (Fig. 5.5, compare lanes 3 and 6, 4 and 7, 5 and 8). An unrelated oligonucleotide did not compete for the complex (lanes 9 to 11). The complex formed by purified Spl protein behaved in an equivalent fashion to complex \(A\) when subjected to competition analysis (compare Figs 5.1 and 5.5). This result is again consistent with the hypothesis that the protein component of complex \(A\) is the murine equivalent of Spl protein.

179
Fig. 5.5. Complex A comigrates with a complex formed between purified Spl protein and the PCMV oligonucleotide. 2.3 ng of the purified protein were used in lanes 2 to 11. Competition experiments with unlabelled oligonucleotides are shown in lanes 3 to 11. Numbers refer to ng of competitor added.
Extract: F9EC Purified HeLa Cell Sp I Protein

Competitor:

PCMV-176/157 MSV-176/157 MSV-363/333

<table>
<thead>
<tr>
<th>ng</th>
<th>100</th>
<th>10</th>
<th>1</th>
<th>100</th>
<th>10</th>
<th>1</th>
<th>100</th>
<th>10</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>A</td>
</tr>
</tbody>
</table>

1 2 3 4 5 6 7 8 9 10 11
Fig. 5.6. A murine equivalent of human Spl protein is detectable by Western blotting. 50 µg of HeLa cell extract or F9 EC cell extract were probed with anti-Spl antibody (anti-Spl) or preimmune serum (P.I). Spl protein is indicated.
Sp1

EXTRACT
HELA
F9 EC

P.I
anti-Sp1
P.I
anti-Sp1

Sp1

Mol. Wt. (Kdal)

110
84
47
33
5.7 Western blot analysis shows that F9 EC cells contain a protein of equivalent size and immuno-reactivity to HeLa cell Spl

Western blot analysis showed that the anti-Spl antibody recognized proteins of equivalent mobility in F9 EC and HeLa cells (Fig. 5.6, compare lanes 2 and 4). The intensity of the signals produced by equivalent concentrations of F9 and HeLa cell protein in the Western blot suggested that F9 EC cells contain a lower concentration of Spl than HeLa cells. Similarly, gel retardation analyses showed that a lower concentration of complex A formed with F9 EC cell extracts than with HeLa cell extracts (Fig. 3.9A, compare lanes 2 and 6). However, it is possible that the anti-Spl antibody has a greater affinity for the HeLa cell Spl than for the F9 cell Spl and therefore the reduced intensity of the signal may not reflect Spl concentration.

The results shown in Figs. 5.3, 5.4, 5.5, and 5.6 are all consistent with the hypothesis that the protein component of complex A is the murine equivalent of human Spl protein. The murine protein has equivalent immuno-reactivity, DNA binding properties and electrophoretic mobility to purified Spl. The protein components of complexes B and C also show equivalent DNA binding properties to Spl, (Fig. 5.1, 5.2). However, they are not recognized by a polyclonal anti-Spl antibody and they have greater electrophoretic mobility than purified Spl in a gel retardation assay. It is possible that these proteins are related to, or derived from, the murine Spl-like protein.

5.8 Wheat germ agglutinin column chromatography suggests that the protein components of complexes A-C are differentially glycosylated

Human Spl protein is heavily glycosylated with O-linked N-acetyl glucosamine (GlcNAc) residues. The presence of these modifications has been shown to play a functional role in the activation of transcription (Jackson and Tjian, 1988). The Spl protein has two major forms of 95kd and 105kd related by phosphorylation (Jackson
Fig. 5.7. Wheat germ agglutinin column chromatography shows that the protein components of complexes B and C are under-glycosylated with respect to the Spl-like protein of complex A. 5.7.A. shows the elution profile of the column. 2ml of wheat germ agglutinin coupled to agarose were packed in a 5ml Bio-Rad column and extensively washed with column buffer (see section 2.23). 7.5mg of F9 EC cell extract were applied to the column, which was then washed at 10 ml/h prior to addition of increasing concentrations of GlcNAc dissolved in column buffer, as indicated in Fig. 5.7.A. Each fraction was of 1.3ml. The elution profile was obtained by using 10μl of each fraction in the Bio-Rad protein concentration assay.

Equivalent protein concentrations (2μg) of each fraction and whole cell extract were then analyzed in a gel retardation assay with the PCMV (Spl WT) probe (Fig. 5.7.B).
A

Protein Concentration (mg/ml)

Fraction No.

0.3 M GlcNAc
0.5 M GlcNAc
0.7 M GlcNAc

B

Fraction

0.3M GlcNAc
0.5M GlcNAc
0.7M GlcNAc

F9EC 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

A  B  C  D
et al., 1990), each of which has multiple sub-forms related by differential glycosylation. Each Spl molecule carries an average of eight glycosylations (Jackson and Tjian, 1988).

The presence of GlcNAc residues may be used to advantage in protein purification. The glycosylated protein selectively binds to the lectin wheat germ agglutinin which can be coupled to agarose and used to form chromatographic columns. The bound protein can then be eluted by addition of free GlcNAc to the column.

The protein components of complexes B and C share some properties with the protein component of complex A, the murine Spl-like protein. The glycosylation properties of the proteins were compared using wheat germ agglutinin column chromatography. F9 EC whole cell extract was applied to a wheat germ agglutinin column, washed extensively, then eluted with increasing concentrations of free GlcNAc. Fractions were collected and assayed by gel retardation.

Fig. 5.7.B shows such a gel retardation analysis, Fig. 5.7.A shows the elution profile of the column and indicates the points at which GlcNAc was added. The flow-through fractions included unbound protein of all three complexes A-C. Addition of 0.3 M GlcNAc did not apparently elute any further protein, but addition of 0.5 M GlcNAc eluted a high concentration of the Spl-like protein which forms complex A. Only low concentrations of complexes B and C were eluted upon addition of GlcNAc. The results of this experiment suggest that the protein components of complexes B and C are under-glycosylated with respect to the Spl-like protein of complex A.

Elution of the Spl-like protein was expected to occur with 0.3 M GlcNAc as described for HeLa cell Spl protein (S. P. Jackson and R. Tjian, pers. comm.). In the experiment described above, and in repeats of this experiment, no significant elution was obtained prior to the addition of 0.5 M GlcNAc. This may be due to the use of slightly different experimental conditions, in particular differences in buffer system. Alternatively, this observation may imply that the F9 EC cell protein is more extensively glycosylated than its HeLa cell equivalent. A comparison of the behaviour of F9 and HeLa cell extracts would be required to test this hypothesis.

187
5.9 HEPARIN-SEPHAROSE ION-EXCHANGE COLUMN CHROMATOGRAPHY SHOWS THAT THE PROTEIN COMPONENT OF COMPLEX B CAN BE SEPARATED FROM THOSE OF COMPLEXES A AND C

Fig. 5.8 shows that the protein component of complex B is less charged than the protein components of complexes A and C. F9 EC whole cell extract was applied to a heparin-Sepharose column and after extensive washing eluted with increasing concentrations of KCl. Fig. 5.8.A shows the elution profile of the column, Fig. 5.8.B shows a gel retardation analysis of alternate fractions collected. Complex B was formed almost exclusively with fraction 26. Complexes A and C were formed predominately with fraction 28 and to a lesser extent with fraction 30. The non-specific complex D formed with all these fractions, and also at low concentrations with all fractions from 18 to 38. Very low concentrations of several other faster mobility complexes were also revealed by the separation.

5.10 DISCUSSION

The experiments described above have demonstrated that three complexes (A, B and C) form at the consensus Sp1 binding site (GGCGGG), around -166 with respect to the transcriptional start in the MPSV and PCMV viral LTRs, with extracts from a variety of cell types. The corresponding wild-type Mo-MSV and Mo-MuLV sequences, which vary by a single base change at -166 from the consensus Sp1 binding site, form equivalent complexes but with at least 5-fold lower affinity.

Polyclonal antibodies raised against human Sp1 interact with the F9 EC cell derived complex A, and purified human Sp1 forms a complex with the PCMV sequence around -166 of equivalent mobility and behaviour in response to competition to complex A. The single point mutation at -166 creates a consensus Sp1 binding site which can interact with a murine Sp1-like protein present in F9 EC cells. This protein has similar immuno-reactivity, DNA binding properties, and electrophoretic mobility to purified human Sp1.

The two faster mobility complexes B and C show similar DNA binding
Fig. 5.8. Heparin-Sepharose column chromatography separates complex B from complexes A and C. 100mg of F9 EC cell protein extract were bound to 10g (dry-weight) of heparin-Sepharose in a 100ml Bio-Rad column and extensively washed with column buffer (section 2.22) at 10ml/h. Protein was then eluted using a linear gradient of 6 mM/min KCl to a final concentration of 1 M KCl. 1.3ml fractions were collected. The elution profile shown in Fig. 5.8.A. was obtained by using 10μl of each fraction in a Bio-Rad protein concentration assay.

Fig. 5.8.B. shows a gel retardation assay using 2μg of protein from alternate fractions in a gel retardation assay with the PCMV (Sp1 WT) probe.
A

B

Probe = PCMV -176/-157

<table>
<thead>
<tr>
<th>Whole Extract</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>46</td>
<td>48</td>
</tr>
</tbody>
</table>
properties to complex A. However, the protein components of these complexes are not recognized by polyclonal anti-Spl antibodies and are under-glycosylated with respect to the Spl-like protein. The identity of these proteins remains unclear but they may be members of a family of Spl-like proteins, or alternatively they may be derivatives of the Spl-like protein. The protein component of complex C is apparently considerably smaller than the other proteins, as estimated from its mobility in gel retardation assays. Complex C is also approximately 5-fold less abundant than the other complexes. These observations suggest that C may be a degradation product of A or B. However, the relative concentrations of the complexes remain approximately equivalent in different extracts, or in the same extract following freezing and thawing, inconsistent with the hypothesis that C forms by degradation during extraction or storage. The under-glycosylation of the protein components of complexes B and C may imply that they have less functional significance than the Spl-like protein. Glycosylation has been shown to increase the activity of Spl in *in vitro* transcription assays (Jackson and Tjian, 1988).

Complexes B and C are not unique to F9 EC cell extracts nor to the PCMV (Spl WT) probe. Fig. 3.9.A shows that a variety of extracts form complexes A-C with the Spl WT probe, including a HeLa cell extract. Oligonucleotide probes corresponding to other characterized Spl binding sites also form an equivalent array of three complexes (E. Spanopoulou, NIMR, London, pers. comm.). The purified Spl used in these experiments was obtained from HeLa cells yet it only forms a complex corresponding to complex A in a gel retardation assay. The HeLa cell Spl was purified by its ability to interact with the GC box in a footprinting assay rather than in a gel retardation assay (Briggs et al., 1986). This may explain why the protein components of complexes B and C were not also purified. Purified HeLa cell Spl protein is fully functional in an *in vitro* transcription assay (Briggs et al., 1986), suggesting that, at least *in vitro*, loss of the faster mobility proteins does not lead to any significant loss of activity.

It is not possible to unequivocally identify the protein components of complexes A-C without further purification. Preliminary
experiments have indicated that such purification would be feasible. Wheat germ agglutinin column chromatography separates the proteins according to glycosylation state and heparin-Sepharose column chromatography separates them according to charge. A combination of these techniques with affinity chromatography should allow each of the three proteins to be isolated. If purified proteins were available protein sequence data could be derived, or one of a number of techniques could be used to isolate cDNA clones for their genes. Only by comparison of amino-acid sequences could the relationships between the three proteins be accurately determined.

The experiments described above have confirmed that the PCMV and MPSV LTR sequences, around -166, form complexes with F9 EC cell proteins of a higher affinity than the complexes formed with the corresponding Mo-MSV or Mo-MuLV LTR sequences. One of these proteins is apparently the murine equivalent of the human transcriptional activator Sp1. The functional significance of these interactions with respect to the increased transcription of the mutant viruses in EC cells is addressed in Chapter 6.
CHAPTER 6

*IN VIVO ANALYSES OF THE ROLE OF POINT MUTATIONS IN THE TRANSCRIPTION OF MUTANT RETROVIRUSES*
6.1 TRANSIENT TRANSFECTION ANALYSES OF THE EFFECTS OF POINT MUTATIONS ON THE TRANSCRIPTION OF Mo-MSV IN EC CELLS

Protein binding sites have been defined by point mutations in the Mo-MSV LTR DNA sequence at -345 and -166 bp upstream of the transcriptional start. The characterization of the DNA-binding proteins interacting with each of these sites has been described in Chapters 4 and 5 respectively. Briefly, the wild-type viral sequence around -345 interacts with an EC cell-specific potential transcriptional repressor, EC-R. The corresponding PCMV sequence has a lower affinity for this factor. Conversely, the wild-type sequence around -166 forms three non-differentiation regulated complexes with a lower affinity than the corresponding MPSV or PCMV sequences. One of these complexes involves a murine equivalent of Spl.

These in vitro results lead to the proposal of the following hypotheses:

1) Introduction of the -166 point mutation into the Mo-MSV LTR, to form a consensus Spl binding site, will allow increased efficiency of transcription in EC cells and possibly in differentiated cells.
2) Introduction of the -345 point mutation into the Mo-MSV LTR, disrupting the EC-R binding site, will allow increased efficiency of transcription only in EC cells.

6.1.1. Introduction of point mutations to pMSVcat.

In order to test these hypotheses the point mutations were introduced separately or together into the pMSVcat construct (see Table 2.5). pMSVcat is a plasmid in which the cat gene coding sequences have been placed under control of the Mo-MSV enhancer and promoter sequences. Transfection of this plasmid into the cell type of interest allows the level of transcription, and hence the activity of the Mo-MSV sequences, to be measured indirectly by a CAT assay. The concentration of CAT enzyme expressed is assumed to be proportional to the level of RNA message transcribed.

The point mutations were introduced into pMSVcat by site-directed mutagenesis using the Bio-Rad muta-gene kit, which has been adapted
Fig. 6.1. Strategy for cloning point mutations into pMSVcat.
1) The XbaI fragment of pMSVcat was excised and subcloned into the polylinker of pT7T318U to form pT7T3(XbaI).
2) The point mutation was introduced at -166 and the XbaI fragment excised and returned to pMSVcat to form pMSV(-166)cat.
3) The point mutation was introduced at -166.
4) The second point mutation was introduced at -345 and the XbaI fragment excised and returned to pMSVcat to form pMSV(-166,-345)cat.
5) The point mutation was introduced at -345 and the XbaI fragment excised and returned to pMSVcat to form pMSV(-345)cat.

■ ■ I  = M o-MSV LTR
[ ]  =  cat gene
[ ]  =  SV40 polyadenylation signal
pMSVcat

①

pT7T318U

② ③ ④ ⑤

pMSV[-166]cat

pMSV[-166,-345]cat

pMSV[-345]cat
from the method of Kunkel (1985), as described in section 2.5. A
single-stranded (ss) DNA template is required for the mutagenesis
reaction and therefore it was necessary to introduce the LTR sequence
into a phagemid vector from which ss phage DNA could be simply
prepared. Fig. 6.1 summarizes the cloning strategy used. An XbaI
fragment from pMSVcat was introduced into the phagemid pT7T318U, to
form pT7T3(XbaI) from which ssDNA containing uracil residues was
derived by growing the phagemid in a dut, ung mutant strain of
bacteria. To produce the -166 and -345 mutations, the negative
strands of the Spl WT and the PCMV oligonucleotides (for sequences
see section 2.2), respectively, were annealed to the ssDNA template
and T4 DNA polymerase and ligase used to synthesize the mutagenic
strand. The covalently closed circular DNA so produced was
transformed into a bacterial strain with active uracil-N-glycosylase
thus selecting for the mutated strand which did not contain uracil.
The double mutant was formed by taking the construct mutated at -166
and repeating the process to introduce the second mutation at -345.
Once mutated the XbaI fragments were excised and returned to pMSVcat
to form pMSV(-166)cat, pMSV(-345)cat, and pMSV(-166,-345)cat. The
synthesis of the -166 mutation is taken as an example and considered
in more detail below.

The concentration of ss pT7T3(XbaI) phagemid DNA was estimated by
electrophoresis through an agarose gel in parallel with the ss
"control phagemid" provided with the muta-gene kit. Reactions were
performed in which 6 or 9pmoles of Spl WT (-) oligonucleotide, 6 or
9pmoles of Spl WT (+) oligonucleotide, or no oligonucleotide, were
annealed to 200ng of ss pT7T3(XbaI) DNA. At the same time reactions
were performed in which 9pmoles of the "reverting primer" or no
oligonucleotide were annealed to 200ng of ss "control phagemid" DNA.
After annealing, T4 DNA polymerase and ligase were used to synthesize
the mutagenic strand. Fig. 6.2 shows electrophoretic analysis of one
tenth of the products from each of these reactions on a 1% (w/v)
agarose gel electrophoresed in TAE buffer containing 0.5 μg/ml
ethidium bromide. In the presence of the "reverting primer" or the
Spl WT (-) oligonucleotide covalently closed circular (ccc) DNA was
Fig. 6.2. Formation of covalently closed circular DNA during formation of the mutagenic strand. Track 1 = 200ng of ssDNA from phagemid pT7T3(XbaI), which includes the Mo-MSV LTR sequences (indicated as Mo-MSV). Tracks 2 to 8 = 1/10 of the products from reactions to synthesize mutagenic strands. The ssDNA used in the reactions was either Mo-MSV or the "control phagemid" provided with the muta-gene kit. The oligonucleotide primers used were the "reverting primer" provided with the muta-gene kit (indicated as control) and the positive and negative strands of the Spl WT oligonucleotide (see section 2.2 for sequence); pM refers to pmoles, s.s. is single-stranded DNA, c.c.c. is covalently closed circular DNA and o.c. is open circular DNA.
detected (tracks 2, 6 and 7).

The remaining products of each of these reactions were transformed into competent MV1190 *E. coli* cells. The control reagents provided with the kit allow the efficiency of mutation to be estimated. A successful mutation of "control phagemid" DNA with the "reversion primer" corrects a point mutation in the *LacZ* gene. In the presence of X-gal and IPTG, as described in section 2.4.4, a blue reaction product will be formed if the mutation has occurred. Table 6.1 summarizes the results of the transformation.

**TABLE 6.1**

Mutagenesis of pMSVcat at -166 and reversion of the "control phagemid".

<table>
<thead>
<tr>
<th>DNA</th>
<th>primer</th>
<th>No. of colonies produced</th>
<th>white</th>
<th>blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. &quot;control&quot;</td>
<td>-</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2. &quot;control&quot;</td>
<td>9pmoles &quot;reverting&quot;</td>
<td></td>
<td>9</td>
<td>32</td>
</tr>
<tr>
<td>3. pT7T3(XbaI)</td>
<td>-</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4. pT7T3(XbaI)</td>
<td>6pmoles Spl WT (+)</td>
<td></td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>5. pT7T3(XbaI)</td>
<td>9pmoles Spl WT (+)</td>
<td></td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>6. pT7T3(XbaI)</td>
<td>6pmoles Spl WT (-)</td>
<td></td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>7. pT7T3(XbaI)</td>
<td>9pmoles Spl WT (-)</td>
<td></td>
<td>220</td>
<td></td>
</tr>
</tbody>
</table>

For each reaction the production of viable colonies corresponded to presence of cccDNA as expected. The control reaction shows that the mutagenesis reaction was 63% efficient.

12 single colonies produced with Spl WT (-) and pT7T3(XbaI) were cultured and plasmid DNA isolated. Each clone was sequenced using the MSV (+) oligonucleotide as a primer and 8/12 shown to carry the mutation at -166. The reaction was therefore approximately 66% efficient. The *XbaI* fragment from one of these positive clones was excised and returned to pMSVcat to form pMSV(-166)cat. The -345 mutation was produced in precisely the same fashion using the MSV (-) oligonucleotide as primer. Fig. 6.3 compares the sequences of
Fig. 6.3. Sequencing of wild-type and mutated sequences around -166. The sequences shown are from the pT7T3(XbaI) plasmid and its mutated derivative.

6.3.A. The upper strand was sequenced from the MSV (+) primer.
6.3.B. The lower strand was sequenced from the T3 primer (Stratagene).
Fig. 6.4. Sequencing of wild-type and mutated sequences around -345. The sequences shown are from the pT7T3(XbaI) plasmid and its mutated derivative.

6.4.A. The upper strand was sequenced from the T7 primer (Stratagene)
6.4.B. The lower strand was sequenced from the Sp1 WT (-) primer.
both the upper and lower strands of pMSVcat and pMSV(-166)cat. Similarly, Fig. 6.4 compares the sequences of pMSVcat and pMSV(-345)cat.

6.1.2. CAT assays.

The commonly used method of performing CAT assays (Gorman et al., 1982b) is time consuming and difficult to quantitate. Briefly, this method requires extracts of transfected cells containing CAT enzyme to be incubated with [14C]-chloramphenicol and acetyl CoA for a set period. Reaction products are then extracted into ethyl acetate and analyzed by thin layer chromatography which separates [14C]-chloramphenicol from acetylated [14C]-chloramphenicol. The reaction products are visualized by autoradiography. Quantitation of results requires scintillation counting of the region of the TLC plate from which each signal is derived.

A more recently developed method of performing CAT assays, known as the "one-vial" method (Sleigh, 1986), allows immediate quantitation of results, produces results at multiple time-points from a single reaction, and allows repeats of individual assays within a single experiment to be carried out. Using this method it is therefore easier and quicker to obtain statistically significant results from CAT assays.

The method relies upon the difference in solubility between [14C]-butyryl CoA and [14C]-butyrylated chloramphenicol in organic solutions. The CAT enzyme is able to catalyze the butyrylation of chloramphenicol as well as its acetylation. Protein extract containing the CAT enzyme is incubated with [14C]-butyryl CoA and unlabelled chloramphenicol in a scintillation vial. These aqueous reagents are overlayed with organic scintillant into which only the reaction products can diffuse. At various time-points the reaction is monitored by scintillation counting; only the 14C in the organic scintillant can be detected and hence the concentration of reaction products is assayed. The reaction can be continued for as long as required allowing multiple readings to be taken and therefore the method is ideal for experiments in which low concentrations of CAT
Fig. 6.5. CAT assay using purified enzyme. Increasing concentrations of CAT enzyme were incubated with 0.1μCi of [14C]-butyryl CoA in 250μl of 1 mM chloramphenicol, 100 mM Tris-HCl (pH 7.8) at room temperature in plastic scintillation vial inserts. The reaction mixtures were overlayed with Econofluor organic scintillant (NEN) and the course of the reaction followed by scintillation counting at 14 min intervals. The graph shows a plot of radioactivity of the reaction products (in counts per minute) with respect to time, after subtraction of the small background value obtained in the absence of enzyme.
Count per minute

20000 n
10000 -
0 100 200

CAT enzyme (units)

• 0.032 U
• 0.016 U
■ 0.008 U
♦ 0.004 U
■ 0.002 U
□ 0.001 U

time / minutes
enzyme are to be assayed.

The linearity of response of the assay, with respect to enzyme concentration and time, was confirmed using commercially available purified CAT enzyme (Sigma). Fig. 6.5 shows a graph in which radioactivity (in counts per minute) is plotted with respect to time for reactions containing various concentrations of CAT enzyme. This graph shows that even at the highest enzyme concentration used the response continues to be linear up to 180 min. The gradient of each line gives a measure of the activity of the enzyme in counts per minute per minute (counts.min^{-3}). As expected, the activity of the enzyme as measured by this assay parallels enzyme concentration, showing that the assay response is linear at the enzyme concentrations tested.

Following the success of the pilot experiment using purified CAT enzyme it was decided to use the "one-vial" method for all CAT assays. In order to minimize potential causes of error during transfection and CAT assays, such as variability in quality of calcium phosphate/DNA coprecipitates, variability in cell numbers between plates of cells, and pipetting errors, all experiments were performed in triplicate. Three plates of cells were used, separately produced precipitates were used to transfect them, and separate CAT assays were performed on the cell extracts.

Another source of error in transfection experiments is variability in the quality of DNA. To minimize such variations plasmids to be used in the same experiment were grown simultaneously, and double-purified using two cesium chloride/ethidium bromide gradients. Prior to use samples of the DNA were electrophoresed through a 0.8% agarose gel to check for degradation. The plasmid DNA was only used if at least 80% was in the supercoiled form. Finally, each experiment was repeated several times using different preparations of DNA.

The efficiency of transfection for the different cell lines used in this study was estimated by transfection with the plasmid pCMVLacZ (see Table 2.5). The cytomegalovirus (CMV) enhancer is highly active in all cell types and hence any cells transfected with this plasmid express the LacZ gene product, 8-galactosidase. This can be simply detected by an assay which produces a blue reaction product (see
The efficiency of tranfection was determined by calculating what proportion of the cells had stained blue. The efficiencies varied by up to 10-fold between experiments but were consistently higher for SV3T3 clone 49 cells (Rigby et al., 1980; an SV40 transformed murine fibroblast line; 1-10%), than for the F9 and PCC4 EC cell lines (0.5-5%).

6.1.3. The Mo-MSV enhancer is active in fibroblasts but inactive in EC cells.

The activity of the Mo-MSV enhancer in differentiated cell types was confirmed by transfection of SV3T3 clone 49 cells. The expression levels from the pMSVcat plasmid, which contains the wild-type enhancer sequence, were compared to expression levels from plasmids pLTR10cat and pLTR0cat (see Table 2.5) which have had the enhancer sequences and the enhancer plus promoter sequences deleted, respectively. Fig. 6.6 shows the results of a CAT assay using extracts of SV3T3 cells transfected with these three constructs. The CAT activity of cells transfected with pLTR0cat was comparable with the activity observed with extracts transfected with carrier DNA only (1.7 counts.min^-2). This value was assumed to be "background" caused by mixing of the aqueous and organic layers at their interface, and was therefore subtracted from the activities obtained with the test constructs. pLTR10cat had low activity (1.0 counts.min^-2) and pMSVcat had 10-fold higher activity (9.9 counts.min^-2). From this result it can be concluded that the Mo-MSV enhancer is active in differentiated cells.

The same three constructs were also used to transfect F9 EC cells. The experiment was repeated four times at DNA concentrations of 1µg, 2.5µg and 5µg for each 50mm diameter plate containing 5x10^6 cells. The results obtained with pLTR0cat and pLTR10cat were very similar to those obtained on transfection of SV3T3 cells. However, extracts of F9 EC cells transfected with the pMSVcat construct, which includes the Mo-MSV enhancer sequences, had similar CAT activity to that produced by the enhancer deleted construct pLTR10cat. The ratio of CAT activity obtained with pMSVcat to that obtained with pLTR10cat...
Fig. 6.6. The Mo-MSV enhancer is active in differentiated cells.
2.5μg of pMSVcat, pLTR10cat, or pLTR0cat were transfected into 5x10^6
SV3T3 clone 49 cells in 50mm diameter dishes. Protein was extracted
from the harvested cells by freeze-thaw lysis and 100μg used in each
assay under the same conditions as described for the control
experiment in Fig. 6.5. The graph shown is a plot of the average
radioactivity from three repeat experiments, after subtraction of the
background value obtained with the promoterless construct pLTR0cat,
with respect to time.
varied between 1.5 and 0.8 and had a mean value of 1.1. The concentration of DNA transfected made no significant difference to this ratio. From these results it may be concluded that the Mo-MSV enhancer is inactive in EC cells, and that its deletion does not affect transcription.

6.1.4. The -166 mutation increases Mo-MSV transcription 6-fold in transient transfection experiments.

The CAT activities of extracts from F9 EC cells transfected with pMSVcat and pMSV(-166)cat were compared in order to assess the functional significance of this mutation. It was predicted that presence of the mutation, creating an Spl binding site, would allow increased transcription. Fig. 6.7 shows this hypothesis to be correct.

Various concentrations of each construct, from 1μg to 10μg, were transfected into 50mm diameter plates of 5x10⁵ F9 EC cells. Extracts of these cells were then used in CAT assays. Fig. 6.7 shows the results of a CAT assay from one such experiment. As the concentration of each test construct transfected was increased the CAT activity increased in direct proportion. At each concentration the activity from the pMSV(-166)cat construct was approximately 6-fold higher than the activity from the pMSVcat construct. This experiment was repeated several times and the results are summarized in Table 6.2. Although the absolute values for CAT activity of each construct varied by up to 10-fold between separate experiments, the ratio of CAT activity between pMSV(-166)cat and pMSVcat always followed the same trend. The mean value for this ratio from all experiments performed was 6, i.e. the -166 mutation increased transcription from the Mo-MSV LTR by 6-fold in F9 EC cells.

The effect of the -166 mutation on transcription in differentiated cells was also investigated. Fig. 6.8 shows the results of a CAT assay using extracts of SV3T3 cells transfected with various concentrations of pMSVcat and pMSV(-166)cat. As with the F9 EC cells CAT activity increased in proportion to the concentration of DNA transfected. At each DNA concentration the CAT activity from the
Fig. 6.7. The -166 mutation increases transcription in F9 EC cells by 6-fold. pMSVcat and pMSV(-166)cat were transfected into F9 EC cells. Extracts of these cells were used in CAT assays. The graphs show plots of the average radioactivity of three repeat experiments, after subtraction of the background activity obtained with pEMBL carrier DNA alone, plotted with respect to time. The concentrations of test construct DNA were as follows: (i) 1µg; (ii) 2.5µg; (iii) 5µg; (iv) 10µg.
TABLE 6.2

Relative CAT expression from pMSV(-166)cat and pMSVcat in F9 EC cells. The results of six separate experiments, using six different DNA preparations are shown. The ratio of pMSV(-166)cat to pMSVcat expression is shown in each case. The CAT activity obtained with carrier DNA alone was deducted from each value prior to calculation of the ratio. Absolute values of CAT activity varied by up to 10-fold between experiments.

<table>
<thead>
<tr>
<th>DNA</th>
<th>Expt.</th>
<th>1 µg</th>
<th>2.5 µg</th>
<th>5 µg</th>
<th>7.5 µg</th>
<th>10 µg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>4.4</td>
<td>10.3</td>
<td>11.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.96</td>
<td>4.35</td>
<td>8.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1.14</td>
<td>1.10</td>
<td>1.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3.10</td>
<td>2.37</td>
<td>2.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>11.8</td>
<td>10.9</td>
<td>18.8</td>
<td>4.40</td>
<td>3.22</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>6.47</td>
<td>4.89</td>
<td>5.69</td>
<td>9.20</td>
<td>7.44</td>
</tr>
</tbody>
</table>
Fig. 6.8. The -166 mutation increases transcription in SV3T3 cells by 2-fold. pMSVcat and pMSV(-166)cat were transfected into SV3T3 clone 49 cells. Extracts of these cells were used in CAT assays. The graphs show plots of the average radioactivity of three repeat experiments, after subtraction of the background activity obtained with pEMBL carry DNA alone, plotted with respect to time. The concentrations of test construct DNA were as follows: (i) 1μg; (ii) 2.5μg; (iii) 5μg.
mutated construct was approximately 2-fold higher than from the wild-type construct, i.e. the -166 mutation increased transcription from the Mo-MSV LTR by 2-fold in a differentiated cell type.

CAT assays do not give information about the transcriptional start-site used. It was considered important to determine the RNA start used by the pMSV(-166)cat construct, as the -166 mutation may increase transcription by activating an upstream start. RNA from SV3T3 cells was analyzed by primer extension from the CAT-1 oligonucleotide. This oligonucleotide hybridizes to sequences within the cat gene coding sequences (for sequence see section 2.2). Fig. 6.9 shows electrophoretic analysis of the extension products, and of sequencing reactions of pMSV(-166)cat also primed with the CAT-1 oligonucleotide. The only extension product co-migrated with the base corresponding to the published transcriptional start (Fuhrman et al., 1981). The reaction was apparently more efficient when the temperature for annealing primer to transcript was 55°C rather than 50°C. The result of this experiment shows that pMSV(-166)cat does use the normal transcriptional start-site in transient transfections of SV3T3 cells. The RNA obtained from transfected F9 EC cells was not at a high enough concentration for the start-site to be analyzed by primer extension. The more sensitive technique of RNAse protection was therefore used. This technique is quantitative and was also used to ensure that CAT enzyme activity reflects mRNA levels. In order to generate an RNA probe for use in this assay a fragment of pMSVcat (from XbaI at -152 bp upstream of the Mo-MSV transcriptional start to PvuII at +158 bp in the cat sequence) was subcloned into the polylinker of pBluescript SK+ (from XbaI to EcoRV) to form plasmid pBS(XP).

A 400 bp continuously [³²P]-labelled riboprobe was synthesized from the T7 promoter of linearized pBS(XP) and used in the RNAse protection assay. Fig. 6.10 shows an RNAse protection experiment using RNA from untransfected F9 EC cells, from F9 EC cells transfected with pMSVcat or pMSV(-166)cat, or from SV3T3 cells transfected with pMSV(-166)cat. The only protected fragment unique to RNA extracted from transfected cells was approximately 194 bp in all cases, which is the expected size for transcripts initiated at the
Fig. 6.9. Primer extension analysis of RNA transcripts from pMSVcat and pMSV(-166)cat in SV3T3 cells. 50μg of RNA isolated from untransfected SV3T3 clone 49 cells or cells transfected with pMSV(-166)cat were analyzed by primer extension from the CAT-1 oligonucleotide (see section 2.2 for sequence). The reaction products were electrophoresed beside sequencing reactions (A-T) from the pMSVcat construct also primed with the CAT-1 oligonucleotide. The temperature for the annealing of primer to RNA was 50°C in tracks 6 and 8 and 55°C in tracks 7 and 9. The major extension product is indicated.
Fig. 6.10. RNAse protection analysis of RNA transcripts from pMSVcat and pMSV(-166)cat in F9 EC cells. 50μg of RNA isolated from untransfected F9 EC cells, F9 EC cells transfected with pMSVcat or pMSV(-166)cat, or from SV3T3 cells transfected with pMSV(-166)cat were analyzed by RNAse protection using a 400 bp riboprobe produced by transcription from the T7 promoter of construct pBS(XP). The probe and the major protected species are indicated.
published start-site. Due to the different electrophoretic properties of RNA and DNA it is not possible to accurately size the protected fragments. However, as the protected fragments obtained with RNA from transfected F9 EC cells co-migrated with the protected fragment from SV3T3 cells it may be assumed that the same transcriptional start was used in both cell types. This result shows that the normal transcriptional start of Mo-MSV is used in F9 EC cells as well as in SV3T3 cells.

Comparison of the signal produced by RNA from F9 EC cells transfected with wild-type and mutant constructs shows that presence of the -166 mutation increased RNA levels by 8-fold. This result is consistent with the results obtained by CAT assays, i.e. CAT activity does indeed reflect RNA levels.

6.1.5. The -345 mutation does not increase Mo-MSV transcription in transient transfection experiments.

The effect of the point mutation at -345 bp on transcription from the Mo-MSV LTR in EC cells was also investigated by transient transfection analyses. *In vitro* experiments showed that this mutation reduced the binding affinity of EC-R, an EC-cell specific potential repressor of transcription, for the LTR. This result leads to the prediction that the -345 mutation may allow increased levels of transcription from the Mo-MSV LTR in EC cells.

pMSVcat and pMSV(-345)cat were transfected into F9 EC cells and the efficiency of their transcription analyzed by CAT assays. The CAT activities of extracts from these cells were very low and approximately equivalent. The CAT activity detected was not significantly greater than the background levels found with extracts transfected with pEMBL. In order to improve the "signal to noise" ratio the efficiency of transcription of pMSV(-166)cat was compared to that of pMSV(-166,-345)cat. Fig. 6.11 shows the result of one such experiment. The CAT activities were significantly greater than background due to the positive effect of the -166 mutation. However, over the range of DNA concentrations used the two constructs were transcribed with approximately equal efficiency. This experiment was
Fig. 6.11. pMSV(-166)cat and pMSV(-166,-345)cat have equivalent transcriptional activity in F9 EC cells. pMSV(-166)cat and pMSV(-166,-345)cat were transfected into F9 EC cells. Extracts of these cells were used in CAT assays. The graphs show plots of the average radioactivity of three repeat experiments, after subtraction of the background activity obtained with pEMBL carrier DNA alone, plotted with respect to time. The concentrations of test construct DNA were as follows: (i) 1μg; (ii) 2.5μg; (iii) 5μg.
Fig. 6.12. Wild-type and mutant Mo-MSV LTR sequences direct equivalent levels of transcription in PCC4 EC cells. 5µg of pMSVcat, pMSV(-166)cat, pMSV(-345)cat and pMSV(-166,-345)cat were transfected into PCC4 EC cells. Extracts of these cells were used in CAT assays. The graphs show plots of the average radioactivity of three repeat experiments, after subtraction of the background activity obtained with pEMBL carrier DNA alone, plotted with respect to time.
Counts per minute

0 - 4000

Time / minutes

Counts per minute vs. time / minutes for:
- pMSVcat
- pMSV(-166)cat
- pMSV(-345)cat
- pMSV(-166,-345)cat
repeated four times and on each occasion a similar result was obtained. The ratio of CAT activity obtained with pMSV(-166)cat to that obtained with pMSV(-166,-345)cat varied between 3.2 and 0.5 with a mean ratio of 1.3. Transient transfection experiments therefore suggest that the -345 mutation does not facilitate any increased efficiency of transcription of Mo-MSV in F9 EC cells.

EC-R, the proposed EC cell repressor of transcription, was detected in both F9 and PCC4 EC cells. The -345 mutation, which decreases the affinity of EC-R for the LTR, is unique to the PCMV mutant virus. Unlike MPSV, which does not have the point mutation at -345, PCMV is efficiently transcribed in both F9 and PCC4 EC cells. MPSV is efficiently transcribed in F9 EC cells only (see Chapter 1, section 1.5.4). These results may indicate that EC-R does not play a significant role in the block to retroviral transcription in F9 EC cells, but does play a significant role in PCC4 EC cells. The comparison of efficiency of transcription of pMSVcat and its mutated derivatives was therefore repeated in PCC4 EC cells.

Fig. 6.12 shows an example of an experiment in which the CAT activities of extracts of PCC4 EC cells transfected with pMSVcat, pMSV(-166)cat, pMSV(-345)cat and pMSV(-166,-345)cat were compared. All four constructs have almost identical expression levels in PCC4 EC cells. This experiment was repeated six times at DNA concentrations between 1μg and 10μg. The mean ratios of CAT activity with pMSVcat to activity with each mutant were as follows:-

\[
\begin{align*}
pMSV(-166)cat &= 1.05 (+/- 0.1) \\
pMSV(-345)cat &= 0.99 (+/- 0.2) \\
pMSV(-166,-345)cat &= 0.98 (+/- 0.4)
\end{align*}
\]

Two important conclusions can be drawn from this transient transfection experiment: -
1) Although the murine Spl-like protein is present in PCC4 EC cell extracts, the -166 point mutation, which creates a consensus Spl binding site, does not increase the efficiency of Mo-MSV transcription in these cells.
2) Despite the ability of the PCMV virus, which has a point mutation at -345, to efficiently transcribe in PCC4 cells, presence of this mutation in Mo-MSV does not allow increased levels of transcription.

228
The transcriptional activities of pMSVcat and pMSV(-345)cat were also compared in differentiated cells. The CAT activities of extracts of SV3T3 clone 49 cells transfected with each construct were very similar. The experiment was performed twice, using 2.5 μg and 5 μg of each construct. The ratio of CAT activity with pMSV(-345)cat to activity with pMSVcat was 1.1 (+/- 0.25).

6.1.6. Discussion.

Two point mutations were introduced into the Mo-MSV LTR sequence in the context of the pMSVcat plasmid using the Bio-Rad muta-gene kit. The -166 mutation corresponds to a point mutation shared by the MPSV and PCMV mutant viruses, both of which have overcome the F9 EC cell block to efficient viral transcription. The -345 mutation corresponds to a point mutation unique to PCMV, which unlike MPSV has also overcome the PCC4 EC cell block to efficient transcription. These mutations were introduced alone or together to produce the constructs pMSV(-166)cat, pMSV(-345)cat and pMSV(-166,-345)cat. Point mutations were generated with an efficiency of greater than 60% using the Bio-Rad muta-gene kit. This high efficiency facilitated identification of the required constructs by dideoxy sequencing. Use of the "one-vial" CAT assay method enabled the efficiency of transcription of these constructs and others to be accurately quantitated following transient transfection.

Comparison of the activity of pMSVcat and the enhancer deleted construct pLTR10cat in SV3T3 clone 49 cells (an SV40 transformed, murine fibroblast cell line), showed that the Mo-MSV enhancer is highly active in differentiated cell types. A similar experiment in F9 EC cells showed that the enhancer has no significant effect on the efficiency of transcription in EC cells.

The in vitro evidence presented in Chapter 4 suggested the hypothesis that presence of the -166 point mutation, creating a consensus SPl binding site, might in part be responsible for the ability of the MPSV and PCMV mutant retroviruses to be efficiently transcribed in EC cells. Comparison of the transcriptional activity of the pMSVcat construct and the pMSV(-166)cat construct showed that
the mutation increased transcription 6-fold in F9 EC cells, consistent with the hypothesis. However, in PCC4 EC cells the mutation had no significant effect on transcription levels.

At first consideration it seems surprising that the -166 mutation does not allow increased transcription levels in PCC4 EC cells, as gel retardation experiments showed that PCC4 and F9 EC cells have equivalent concentrations of the murine Spl-like protein. However, the mechanism of transcriptional activation by Spl has recently been shown to involve a non DNA-binding coactivator (Pugh and Tjian, 1990). Absence of the coactivator may prevent transcriptional activation via Spl in PCC4 EC cells. Alternatively, PCC4 EC cells may contain factors, not present in F9 EC cells, which repress transcription of Mo-MSV in a manner that cannot be overridden by activators of transcription. The inability of the Spl site created by the -166 mutation to activate transcription in PCC4 EC cells may explain the inactivity of the MPSV mutant retrovirus in this cell type. Further mutations are required, as found in PCMV, to allow efficient transcription to occur in PCC4 EC cells. One candidate mutation responsible for allowing efficient transcription of PCMV is the -345 mutation.

The in vitro evidence presented in Chapter 5 suggested that the EC-R protein, which interacts with the Mo-MSV LTR around -345 bp, may be an EC cell repressor of transcription. The mutation at -345 in PCMV, which reduces the binding affinity of EC-R for the LTR, was therefore hypothesized to play a role in allowing the efficient EC cell transcription of PCMV. However, comparison of the transcriptional activity of the pMSV(-166)cat and pMSV(-166,-345)cat constructs in F9 EC cells showed that the -345 mutation had no effect in this cell type, suggesting that the hypothesis is incorrect.

The -345 mutation does not occur in MPSV, and is therefore not a requirement for efficient transcription in F9 EC cells. The experiment was repeated in PCC4 EC cells, in which MPSV cannot be transcribed but PCMV can be transcribed. Once again transient transfection experiments show no effect of the -345 mutation on transcriptional activity. These results are inconsistent with the hypothesis that EC-R is a repressor of transcription which prevents
Mo-MSV transcription in EC cells.

6.2 REGULATION OF RETROVIRUS TRANSCRIPTION BY THE ADENOVIRUS E1A GENE PRODUCT

Several groups have presented evidence suggesting that F9 EC cells contain an E1A-like activity, this work is reviewed in section 1.4.4. The transcription of papovaviruses is repressed by E1A in an enhancer dependent fashion (Velcich and Ziff, 1985). However, the effects of the E1A gene product on retrovirus transcription have not previously been investigated. The work of Hen et al. (1986) suggests that the transcriptional inactivity of the papovaviruses in EC cells is a function of the E1A-like factor. The block to EC cell transcription of retroviruses may be mediated by a similar mechanism. This suggestion leads to the hypothesis that retroviral transcription may also be subject to repression by the E1A gene product.

The polyomavirus F9.1 mutation is not subject to either EC cell or E1A repression of transcription (Hen et al., 1986). Were Mo-MSV to be repressed by the E1A gene product it would be interesting to determine whether presence of the -166 or -345 point mutations, occurring in mutant viruses capable of expression in EC cells, affect this repression.

To investigate whether the E1A gene product represses transcription of Mo-MSV co-transfection experiments were performed. Various concentrations of the plasmid pRSV₆E1A (see Table 2.5), which expresses both the 12S and 13S RNAs of Ad5 from the RSV enhancer and promoter, were co-transfected into SV3T3 clone 49 cells with pMSVcat. The effect of the presence of E1A gene products on CAT expression from the Mo-MSV LTR was then analyzed by CAT assays. In order to control for promoter competition between the RSV and Mo-MSV sequences a control plasmid, pRSVLacZ, was also co-transfected with pMSVcat. pRSVLacZ includes the same promoter and enhancer sequences as pRSV₆E1A, but expresses β-galactosidase.
6.2.1. The adenovirus E1A gene product down-regulates transcription from the Mo-MSV LTR.

It was considered important to confirm that the E1A gene products were efficiently expressed from the pRSV₆E1A plasmid in SV3T3 clone 49 cells. Western blot analysis of extracts of transfected cells was therefore carried out. The blot was probed with a polyclonal antibody raised against Ad5 E1A products (M73; a kind gift from N. Jones). Fig. 6.13 shows that the antibody recognized a cluster of several proteins in the extract of cells transfected with pRSV₆E1A. No proteins were detected in extracts of untransfected cells or cells transfected with pRSVlacZ. Comparison with the molecular weight standards shows the proteins to range from approximately 40 to 55kd, consistent with the results of Yee et al. (1983). Similarly, to confirm that the pRSVlacZ plasmid was efficiently expressed, transfected cells were assayed for β-galactosidase expression (as described in section 2.14). Blue stained cells expressing β-galactosidase were only found in cells transfected with pRSVlacZ.

The results of CAT assays using extracts of SV3T3 clone 49 cells co-transfected with pRSV₆E1A and pMSVcat, or pRSVlacZ and pMSVcat are shown in Table 6.3. The addition of increasing concentrations of pRSVlacZ causes a small reduction in CAT activity from the Mo-MSV LTR, presumably due to competition for transcription factors required by both RSV and Mo-MSV transcriptional control sequences. However, the addition of increasing concentrations of pRSV₆E1A causes a much larger reduction in CAT activity from the Mo-MSV LTR. This result shows that the E1A gene product does indeed cause repression of transcription from the Mo-MSV LTR.

The repression of transcription of SV40, polyoma and the immunoglobulin kappa light chain gene by E1A have all been shown to be enhancer dependent (Borrelli et al., 1985; Hen et al., 1985; Velcich and Ziff, 1985). To determine whether E1A repression of Mo-MSV transcription is also enhancer dependent the co-transfection experiment was repeated using the enhancer deleted construct pLTR10cat (see Table 2.5) instead of pMSVcat. The results of this experiment are shown in Table 6.4. The absolute CAT activity of
Fig. 6.13. Western blot analysis shows that E1A protein is expressed from the pRSV₆E1A expression plasmid on transfection of SV3T3 cells. 100μg of protein extract from untransfected SV3T3 clone 49 cells, cells transfected with pRSVLacZ, or cells transfected with pRSV₆E1A were electrophoresed through a 7.5% SDS-polyacrylamide gel. The protein was transferred to nitrocellulose and probed with a 1/100 dilution of ascites fluid containing the polyclonal antibody M73 (raised against Ad5 E1A gene products). Size markers were the Bio-Rad prestained low range molecular weight standards.
Expression from pMSVcat is down-regulated by the E1A gene products. In each transfection experiment 2.5µg of pMSVcat were transfected into 5x10⁶ SV3T3 clone 49 cells. The concentration of co-transfected plasmid is indicated. CAT activity is expressed in counts.min⁻², after subtraction of the background CAT activity obtained with transfection of pEMBL, the control DNA.

<table>
<thead>
<tr>
<th>Expt.</th>
<th>Co-transfected construct (µg)</th>
<th>CAT activity</th>
<th>% of CAT activity with pMSVcat alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>-</td>
<td>5.18</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>pRSV₅E1A (2.5)</td>
<td>1.72</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>pRSV₅E1A (5.0)</td>
<td>0.73</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>pRSV₅E1A (7.5)</td>
<td>0.16</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>pRSVLacZ (2.5)</td>
<td>3.48</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>pRSVLacZ (5.0)</td>
<td>2.45</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>pRSVLacZ (7.5)</td>
<td>7.04</td>
<td>136</td>
</tr>
<tr>
<td>2.</td>
<td>-</td>
<td>2.64</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>pRSV₅E1A (1.0)</td>
<td>0.91</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>pRSV₅E1A (2.5)</td>
<td>0.70</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>pRSV₅E1A (5.0)</td>
<td>0.41</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>pRSVLacZ (1.0)</td>
<td>2.17</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>pRSVLacZ (2.5)</td>
<td>1.33</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>pRSVLacZ (5.0)</td>
<td>1.33</td>
<td>50</td>
</tr>
<tr>
<td>3.</td>
<td>-</td>
<td>5.59</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>pRSV₅E1A (1.0)</td>
<td>3.03</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>pRSV₅E1A (2.5)</td>
<td>1.31</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>pRSV₅E1A (5.0)</td>
<td>1.10</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>pRSVLacZ (1.0)</td>
<td>8.37</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>pRSVLacZ (2.5)</td>
<td>3.57</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>pRSVLacZ (5.0)</td>
<td>4.62</td>
<td>83</td>
</tr>
</tbody>
</table>
TABLE 6.4

E1A down-regulation of Mo-MSV transcription is enhancer dependent and point-mutation independent. In each experiment the indicated concentrations of test plasmid and co-transfected plasmid were transfected into 5x10⁵ SV3T3 clone 49 cells. CAT activity is expressed in counts.min⁻², after subtraction of the background CAT activity obtained with transfection of pEMBL, the control DNA.

<table>
<thead>
<tr>
<th>Test construct (μg)</th>
<th>Co-transfected construct (μg)</th>
<th>CAT activity</th>
<th>% of CAT activity with test construct alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pMSVcat (5)</td>
<td></td>
<td>2.20</td>
<td>100</td>
</tr>
<tr>
<td>pMSVcat (5)</td>
<td>pRSV₅E1A (5)</td>
<td>0.55</td>
<td>25</td>
</tr>
<tr>
<td>pMSVcat (5)</td>
<td>pRSVLacZ (5)</td>
<td>1.20</td>
<td>55</td>
</tr>
<tr>
<td>pLTR10cat (5)</td>
<td></td>
<td>1.18</td>
<td>100</td>
</tr>
<tr>
<td>pLTR10cat (5)</td>
<td>pRSV₅E1A (5)</td>
<td>0.85</td>
<td>72</td>
</tr>
<tr>
<td>pLTR10cat (5)</td>
<td>pRSVLacZ (5)</td>
<td>0.43</td>
<td>36</td>
</tr>
<tr>
<td>Experiment 2.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pMSV(-166)cat (2.5)</td>
<td></td>
<td>4.92</td>
<td>100</td>
</tr>
<tr>
<td>pMSV(-166)cat (2.5)</td>
<td>pRSV₅E1A (2.5)</td>
<td>1.43</td>
<td>29</td>
</tr>
<tr>
<td>pMSV(-166)cat (2.5)</td>
<td>pRSV₅E1A (5.0)</td>
<td>0.85</td>
<td>17</td>
</tr>
<tr>
<td>pMSV(-166)cat (2.5)</td>
<td>pRSVLacZ (2.5)</td>
<td>3.68</td>
<td>75</td>
</tr>
<tr>
<td>pMSV(-166)cat (2.5)</td>
<td>pRSVLacZ (5.0)</td>
<td>1.72</td>
<td>35</td>
</tr>
<tr>
<td>Experiment 3.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pMSV(-345)cat (2.5)</td>
<td></td>
<td>14.23</td>
<td>100</td>
</tr>
<tr>
<td>pMSV(-345)cat (2.5)</td>
<td>pRSV₅E1A (1.0)</td>
<td>13.67</td>
<td>96</td>
</tr>
<tr>
<td>pMSV(-345)cat (2.5)</td>
<td>pRSV₅E1A (2.5)</td>
<td>3.76</td>
<td>26</td>
</tr>
<tr>
<td>pMSV(-345)cat (2.5)</td>
<td>pRSV₅E1A (5.0)</td>
<td>1.76</td>
<td>12</td>
</tr>
<tr>
<td>pMSV(-345)cat (2.5)</td>
<td>pRSVLacZ (1.0)</td>
<td>17.07</td>
<td>119</td>
</tr>
<tr>
<td>pMSV(-345)cat (2.5)</td>
<td>pRSVLacZ (2.5)</td>
<td>13.07</td>
<td>92</td>
</tr>
<tr>
<td>pMSV(-345)cat (2.5)</td>
<td>pRSVLacZ (5.0)</td>
<td>8.9</td>
<td>61</td>
</tr>
</tbody>
</table>
pLTR10cat was considerably lower than that of pMSVcat as expected. Co-transfection of an equivalent concentration of pRSV5E1A had little affect on this low level of CAT activity; co-transfection of the control plasmid caused a greater reduction. This result confirms that E1A repression of Mo-MSV transcription is enhancer dependent.

The experiment was also repeated using the pMSV(-166)cat and pMSV(-345)cat constructs in place of the wild-type pMSVcat plasmid (see Table 6.4). The results of these experiments are very similar to those obtained using a wild-type Mo-MSV LTR, they show that presence of the -166 or -345 point mutations in the Mo-MSV LTR does not prevent the E1A gene product from repressing transcription.

The results of the co-transfection experiments described above are summarized in Fig. 6.14 in the form of a histogram. The reduction in CAT activity of pMSVcat, pMSV(-166)cat, pMSV(-345)cat and the enhancer deleted construct pLTR10cat on co-transfection with an equivalent concentration of pRSV5E1A or pRSVLacZ is illustrated. CAT activity from the constructs containing wild-type or point mutated enhancer sequences was reduced 3-fold to 4-fold in the presence of the E1A gene products. However, CAT activity from the enhancer deleted construct was only reduced approximately 1.3-fold by the E1A gene products.

6.2.2. Discussion.

The results of co-transfection experiments show that the E1A gene products repress transcription of wild-type Mo-MSV, and two mutant derivatives, in an enhancer dependent fashion. The plasmid used to express E1A produces both the 12S and 13S mRNAs, which in turn produce both the 243 amino acid and 289 amino acid proteins responsible for repression and activation of transcription respectively. The repression of retroviral transcription by the E1A gene products may therefore be mediated by one of several different mechanisms. The 289 amino acid activator of transcription could cause repression by "squelching", i.e. by sequestering other activators of transcription in solution. Alternatively, the 243 amino acid protein
Fig. 6.14. E1A down-regulates transcription from the Mo-MSV LTR in an enhancer dependent, and point mutation independent fashion. SV3T3 clone 49 cells were transfected with the following test constructs pMSVcat (1), pMSV(-166)cat (2), pMSV(-345)cat (3) and pLTR10cat (4). Co-transfection experiments with an equivalent concentration of RSVLacZ or RSV-E1A were also performed. The CAT activity of cell extracts is represented as a percentage of the CAT activity obtained when only pEMBL carrier DNA was co-tranfected with each test construct.
might cause the repression e.g. by inactivating other transcription factors which interact with the enhancer.

These results parallel those obtained with papovaviruses, i.e. both retroviruses and papovaviruses are inefficiently transcribed in EC cells, and both are repressed by the adenovirus E1A gene products. This observation suggests that the control sequences of these different viruses may share common properties. However, it is possible that both the E1A gene product and the "E1A-like" factor postulated to be present in EC cells have the capacity to interact with a number of different enhancer sequences. Such an interaction may be mediated by a non-DNA binding protein involved in transmitting the effect of DNA binding activators of transcription to the transcriptional start.
CHAPTER 7

FINAL DISCUSSION
7.1 SUMMARY OF CONCLUSIONS

The major conclusions of the results presented in Chapters 3-6 are summarized below.

1) An investigation of DNA-binding proteins interacting with the LTR sequences of Mo-MSV, and the mutant retroviruses MPSV and PCMV, was carried out using the gel retardation assay. Comparison of the binding proteins from embryonal carcinoma stem cells and their differentiated derivatives revealed the following:
   a) Equivalent complexes form with either undifferentiated or differentiated cell extracts on the Mo-MSV LTR between -252 and -152.
   b) An EC cell specific complex forms with the Mo-MSV LTR between -372 and -280.
   c) An EC cell specific complex forms with the Mo-MSV sequence between -363 and -333. This complex can also form on the corresponding PCMV sequence, which has a single point mutation at -345, but with a 6-fold lower affinity.
   d) A parietal-endoderm specific complex forms with equal efficiency on the Mo-MSV and PCMV sequences between -363 and -333.
   e) Three non-differentiation regulated complexes form with the MPSV (or PCMV) sequence from -176 to -157. The same complexes form with the corresponding Mo-MSV sequence but with a 5-fold lower affinity.
   f) The Mo-MSV and the corresponding mutant sequences from -333 to -309 form equivalent arrays of complexes with EC cell extracts. Similarly, these two probes form equivalent complexes with extracts from differentiated cells. However, there are some qualitative differences between the complexes that form with the EC cell and differentiated cell extracts.

2) The -372 to -280 restriction fragment probe includes the -363 to -333 oligonucleotide probe. The EC cell-specific complexes which formed with each of these probes had similar mobility and potentially involved the same protein. The protein had a higher affinity for the wild-type viral sequence, which is transcriptionally inactive in EC cells, and is EC cell-specific. These results suggested that the
protein is a repressor of transcription and therefore it has been termed EC-R for EC cell-Repressor. A further study of this protein was undertaken, revealing the following:

a) The oligonucleotide probe efficiently competes for the EC cell-specific complex that forms with the restriction fragment. This result is consistent with the idea that the same protein complexes with each of the probes.
b) Off-rate analysis shows the complex to have 6-fold higher affinity for the wild-type sequence than for the mutant sequence.
c) The proposed binding site for the EC cell-specific protein has some homology with the binding site for the transcription factor PEA-2. However, gel retardation experiments using a known PEA-2 binding site sequence as a competitor suggest that EC-R is not PEA-2.
d) A computer search for sequences homologous to the EC-R binding site shows that the sequence occurs upstream of several murine cellular genes. The sequence is also conserved between several different retroviral LTR sequences.
e) EC-R has a native molecular weight of approximately 60kd, and is non-glycosylated.
f) Heparin-Sepharose column chromatography purifies EC-R 20-fold. A single round of oligonucleotide affinity chromatography affords a further 10-fold purification of EC-R.

3) The -166 mutation creates a consensus binding site for the transcription factor Sp1. Three non-differentiation regulated complexes (A-C) formed with the PCMV (or MPSV) sequence around -166. The same complexes formed, but with a lower affinity, with the corresponding wild-type sequence. These results led to the hypothesis that the mutant sequences are able to interact more efficiently than the wild-type sequence with an activator of transcription; the activator could be related to or equivalent to Sp1. Further investigations of these complexes revealed the following:

a) Complexes A-C are all efficiently competed by a restriction fragment from the HSV tk promoter, which includes two known binding sites for Sp1.
b) A gel retardation assay using a titration of competitor DNA
concentrations shows that complexes A-C form with 5-fold higher affinity on the mutant sequences around -166.

c) Complex A from murine F9 EC cells is recognized by an antibody raised against human Spl protein. The antibody interacts with A to form a ternary complex, A’.
d) Purified Spl from HeLa cells (a human epithelial cell line) has equivalent electrophoretic mobility to complex A in a gel retardation experiment. The purified Spl also behaves like complex A in competition experiments.
e) Wheat germ agglutinin column chromatography shows that the protein components of complexes B and C are under-glycosylated with respect to complex A.

4) As the -345 and -166 point mutations had both been shown to affect protein-DNA interactions in vitro, it was decided to investigate their roles in the control of transcription using transient transfection analyses.
a) Mutant derivatives of the pMSVcat construct, in which the Mo-MSV LTR sequences drive transcription of the marker gene cat, were successfully formed using site-directed mutagenesis.
b) The Mo-MSV enhancer sequences increase transcription in SV3T3 clone 49 cells (a transformed, murine fibroblast cell line) by 10-fold.
c) The enhancer sequences have no effect on transcription in F9 EC cells.
d) The -166 point mutation increases transcription by 6-fold in F9 EC cells. The usual transcriptional start is used by this mutant construct.
e) The -345 point mutation has no effect on transcription in F9 EC cells.
f) Neither the -166 nor the -345 point mutations affect transcription in PCC4 EC cells.
g) The -166 point mutation increases transcription by 2-fold in SV3T3 clone 49 cells.

5) The gene products of the adenovirus early gene E1A are known to
down-regulate papovavirus transcription. Like retroviruses, the papovaviruses are inefficiently transcribed in EC cells. However, the effect of the E1A gene products on retroviral transcription has not previously been investigated. Co-transfection experiments were performed in order to investigate whether E1A down-regulates retroviral transcription, and whether the -166 and -345 point mutations might modulate such an effect. These experiments revealed the following:

a) E1A down-regulates transcription from the Mo-MSV LTR in SV3T3 clone 49 cells.
b) The down-regulation is dependent upon presence of the enhancer sequences.
c) The -166 and -345 point mutations do not affect this down-regulation of transcription.

In summary, the MPSV and PCMV mutant retroviruses have gained a functional Sp1 binding site which allows increased transcription in F9 EC cells. PCMV has lost an efficient binding site for EC-R, however, this mutation does not allow increased transcription in EC cells in a transient transfection assay. E1A represses transcription from the Mo-MSV LTR in an enhancer dependent, and point mutation independent fashion. These conclusions are now considered in broader terms, in the light of previously published work.

7.2 FURTHER CONSIDERATION OF THE ANALYSIS OF LTR BINDING PROTEINS

The proteins interacting with the Mo-MuLV LTR sequences have previously been investigated by Speck and Baltimore (1987). The binding sites for six distinct nuclear factors from extracts of a mouse B-cell line, WEHI 231, were identified by a combination of gel retardation and footprinting assays. Within each direct repeat sequence of the LTR, binding sites with homology to the NF-1, GRE, and core binding sites were identified. Two NF-1 sites occur in each direct repeat, and a third GRE site occurs upstream of the direct repeat sequences. Three previously unknown binding sites LVa, LVb and LVc were also found. Gel retardation experiments suggested that the
proteins interacting with the core and NF-1 sites are at low
countentration in F9 and PCC4 EC cell extracts, in comparison to
extracts of WEHI 231 cells and NIH 3T3 (murine fibroblast) cells.

More recent work (Speck et al., 1990a) investigated the effect on
efficiency of transcription of mutating the known binding sites. In
the fibroblastic NIH 3T3 cell line, mutations abolishing binding at
the core sequence did not significantly alter the efficiency of
transcription. This result suggests that the lack of core binding
proteins in EC cells is not responsible for the inefficiency of
retrovirus transcription in these cells; however, in the absence of
other activators of transcription the concentrations of core site
binding proteins may become more critical. Mutation of the promoter
proximal NF-1 sites in each direct repeat caused a dramatic reduction
in transcriptional activity in NIH 3T3 cells, to 4% of the wild-type
level. This result implies that lack of NF-1 site binding proteins in
EC cells may have a significant role in preventing efficient
retroviral transcription.

Mutation of the LVa sites increased transcription to 177% of the
wild-type level, mutation of the LVb sites decreased transcription to
30% of the wild-type level, and mutation of the LVc sites increased
transcription to 150% of the wild-type level in NIH 3T3 cells.
Finally, mutation of the two most promoter-distal GRE sites reduced
transcription to 27% of the wild-type level. As no glucocorticoid
hormone was present in these experiments, and transcriptional
activation by the glucocorticoid receptor does not occur in the
absence of ligand, Speck et al. (1990a) infer that transcriptional
activation is a function of other proteins interacting at the GRE
sites.

In this study a restriction fragment from -252 to -152 was used in
gel retardation experiments. This fragment includes the LVb, LVc,
GRE, NF-1 and core binding sites characterized by Speck and Baltimore
(1987). However, no differentiation regulated complexes formed with
this probe. This result does not agree with the findings of Speck and
Baltimore (1987; discussed above) which suggest that core and NF-1
site binding proteins are present only at relatively low
concentrations in EC cells. One possible explanation for this anomaly
is that parietal endoderm-like cells, the differentiated derivatives of F9 cells, are also deficient in these factors. However, a very similar result was obtained using an extract from HeLa cells (a human epithelial cell line). A second potential explanation is that the differences between the results reported here and the results of Speck et al. (1990a) may be a function of the different experimental conditions used. In this study extracts of whole cells rather than nuclei, and different buffer and competitor DNA, were used for gel retardation assays. Finally, the Mo-MuLV and Mo-MSV sequences differ within the NF-1 binding site; Mo-MuLV, which was used in the study by Speck and Baltimore (1987) has a single extra base-pair inserted between -245 and -246 of the Mo-MSV sequence, which was used in this study. This insertion may influence binding at the NF-1 site, within which it lies. The results of this study show that no differentiation regulated binding proteins interact with this fragment, suggesting that sites within this fragment are not responsible for the inefficient transcription of Mo-MSV in EC cells.

The NF-1 WT oligonucleotide used in the gel retardation experiments described in this study centres around the NF-1 binding site, but also overlaps with both the LVa and LVb binding sites. This may explain why multiple complexes form with this probe, unlike the -252 to -152 probe. Some of these complexes are apparently differentiation regulated. However, the mutation at -327, which lies between the binding sites for NF-1 and LVa, does not seem to alter the efficiency of complex formation. This result suggests that the -327 mutation does not play a significant role in allowing efficient transcription of mutant retroviruses in EC cells. The gel retardation technique is very sensitive, but complexes with low affinity may be undetectable due to the presence of much higher affinity complexes with similar electrophoretic mobility. Such undetectable, weak interactions may nevertheless be involved in repressing transcription via the Mo-MSV LTR or activating transcription via the mutant LTRs.

The two restriction fragment probes used in gel retardation assays cover between them all the sequence within a single direct repeat of the LTR. One direct repeat sequence has been lost in the mutant PCMV which can express in PCC4 EC cells. It might therefore be predicted
that an EC cell-specific repressor of transcription interacts with the direct repeat sequence. Such a protein was not detected by gel retardation analyses. The deletion in PCMV may in fact have been an unselected, fortuitous event, which occurred by homologous recombination.

7.2.1. The -345 bp binding proteins.

An EC cell-specific complex, b, was found to form with the -372 to -280 bp fragment of the Mo-MSV LTR. No other complexes specific to the probe were detected. This is surprising as the fragment includes the sequence of the NF-1 oligonucleotide which forms several different complexes, none of which appear to correspond to complex b. Binding to other sites within the fragment could preclude binding to the NF-1 site. Alternatively, the multiple complexes observed with the NF-1 oligonucleotide may be an artefact of using a short isolated stretch of DNA, removed from its usual flanking sequences.

Complex b is efficiently competed by a fragment from -372 to -329 bp, which does not include the majority of the NF-1 oligonucleotide sequence, confirming that b does not correspond to one of the complexes formed with the NF-1 site. Complex b is also efficiently competed by an oligonucleotide corresponding to the -363 to -333 sequence of the Mo-MSV LTR. This oligonucleotide forms an EC cell-specific complex of similar mobility to b, strongly suggesting that the binding site of b is localized to this region of the LTR. This result was confirmed during the course of this study by the results of Tsukiyama et al. (1989). This group also identified an EC cell-specific protein, occurring in PCC4, PCC3 and F9 EC cells, which interacts with the Mo-MuLV LTR. Using nuclear extracts this group were able to further localize the binding site of the protein, which they term ELP (for embryonic LTR binding protein), by the dimethyl sulphate protection assay. The binding site was defined as involving a sequence identical to the Mo-MSV LTR sequence from -351 to -345.

The PCMV sequence from -363 to -333 includes a point mutation at -345. This mutation lowers the efficiency of formation of complex b. This data was also confirmed by Tsukiyama et al. (1989), who note
that -345 is the site of the point mutation in PCMV and state that the mutated sequence has a lower binding efficiency for ELP. The protein component of complex b is obviously a strong candidate for a repressor of retrovirus transcription, as it only occurs in those cell types in which transcription is inefficient, and does not efficiently bind to a mutant virus which is efficiently transcribed. The protein has therefore been termed EC-R for EC cell-Repressor.

The results of this study suggest that EC-R is present at lower concentrations in PCC4 cells than in F9 cells. However, Tsukiyama et al. (1990) did not report any such variation. It should be noted that several different preparations of F9 EC cell extract were tested, but only a single preparation of PCC4 cell extract. Further analysis is therefore required before it can be positively concluded that EC-R concentrations vary between EC cell types.

A complex specific to parietal endoderm-like cells also formed with the Mo-MSV -363 to -333 oligonucleotide. This complex, c, formed with an equivalent affinity with the PCMV oligonucleotide and is therefore likely to have a different binding site specificity to EC-R. The protein may be an activator of transcription, involved in mediating the high transcriptional efficiency of Mo-MSV in differentiated F9 EC cells. This complex was not identified when using the -372 to -280 restriction fragment probe. This discrepancy may reflect the lower resolution achievable with a larger probe. It is possible that complex c is masked by the non-specific complex which forms with the restriction fragment probe.

7.2.2. The -166 bp binding proteins.

The results of gel retardation experiments using the PCMV sequence around -166 (equivalent to the MPSV sequence) strongly suggest that a murine homologue of human Sp1 protein is interacting with this site. During the preparation of this thesis Darrow et al. (1990) published supporting evidence for this hypothesis. This group investigated the control of transcription of the tissue plasminogen activator (t-PA) gene during F9 EC cell differentiation. The t-PA gene promoter includes two GC box motifs and a probe including these sequences was
used in gel retardation experiments. Four complexes formed with extracts from F9 EC cells or their differentiated derivatives; differentiation did not result in any qualitative or quantitative changes to the gel retardation pattern observed. Darrow et al. (1990) showed that the complexes are efficiently competed by the SV40 21 bp repeat sequences, which are known to bind Spl. However, the complexes were not competed by a binding site for AP-2 which also interacts with the 21 bp repeats.

Mutation of one of the two GC boxes, to leave a single consensus Spl binding site analogous to the site present in the PCMV LTR, reduced the number of retarded complexes to two, however in some of the data shown a third complex is visible. These three complexes apparently correspond to complexes A-C as described in this study. The variable resolution of complex B in the experiments of Darrow et al. (1990) may be a consequence of the different buffer system used for electrophoresis (TBE rather than TAE, see Table 2.1), or the different reaction conditions used.

In further agreement with the results presented in this study, Darrow et al. (1990) showed that 2892-E anti-Spl antibody interacts with the Spl-like complexes which form on the t-PA promoter. However, they found that in addition to formation of a slow mobility ternary complex, the antibody also competes for complex formation when at high concentrations. Formation of the fastest mobility complex, which apparently corresponds to the complex identified as C in this study, is not competed. The slower mobility complexes identified by Darrow et al. (1990) require two Spl binding sites and were therefore not seen with the PCMV probe. These complexes may involve the same protein as complex A; this assumption reconciles the differences between the two sets of data, as complex A was also found to be depleted on formation of the ternary complex in this study.

Darrow et al. (1990) also performed gel retardation experiments using an N-terminal deleted form of Spl expressed in E. coli and purified by affinity chromatography (Spl-168C). Spl-168C formed multiple complexes with the t-PA promoter probe; these complexes were of faster mobility than those formed using F9 cell extracts, presumably reflecting the truncation of the protein. Unfortunately,
it is not possible to directly compare these results with the results presented in this study. Firstly, due to use of a gel retardation probe including two Spl binding sites rather than one, and secondly, because truncated protein produced in *E. coli* rather than HeLa cell protein was used. However, both studies show that purified Spl protein interacts with Spl sites with a similar binding specificity to the murine protein.

The results presented in this study showed that complex A was not only recognized by the anti-Spl antibody, but also has identical electrophoretic mobility and behaviour in competition experiments to the purified HeLa cell Spl protein. These results strongly suggest that complex A involves a murine homologue of Spl protein. However, the identity of the protein components of complexes B and C is more difficult to assess. These proteins could be derivatives or relatives of murine Spl, or they could be unrelated to Spl.

In favour of the former hypothesis is the similar binding specificity of the different proteins; they apparently respond identically to competition with the wild-type and point mutated GC box sequences. However, these proteins are not recognized by the anti-Spl antibody. The antibody was raised against an *E. coli* synthesized fusion protein lacking the N-terminal 20% of Spl. As complexes B and C have a faster mobility than complex A it is likely that they involve smaller proteins, these could be early termination products of full-length Spl protein, and hence be predominantly N-terminal in origin. Alternatively, these proteins could be folded or post-translationally modified in such a way that the Spl antibody no longer recognizes them. Should these proteins represent part of a "family" of Spl-like proteins they may be too far diverged from Spl protein itself to be immunologically related.

Wheat germ agglutinin (WGA) column chromatography showed that the protein components of complexes B and C are under-glycosylated with respect to complex A. Glycosylation is required for full transcription activator activity of Spl (Jackson and Tjian, 1988). Hence, the less glycosylated proteins may not have a functional role *in vivo*. The purified HeLa cell Spl protein, which formed a complex with the PCMV GC box of equivalenty mobility to complex A, was
purified by WGA column chromatography and affinity chromatography, and its activity assessed by footprinting analysis rather than by gel retardation. This purification method may have led to loss of the HeLa cell proteins that form complexes B and C. The gel retardation experiments performed in this study using extracts from HeLa cells clearly show that complexes form with human proteins of equivalent electrophoretic mobility to B and C. Formation of complexes B and C does not seem to be a function of the flanking sequences surrounding the GC box in the PCMV LTR as a similar array of complexes has been observed with GC box probes from the Thy-1 promoter (E. Spanopoulou, NIMR, London, pers. comm.).

The hypothesis that the protein components of complexes B and C are not related to Spl protein is apparently supported by the methylation interference analyses of Darrow et al. (1990). The different complexes which form on the t-PA GC boxes have different contact points with the DNA, although once again it is not possible to directly correlate the results of Darrow et al. (1990) with the results presented in this study due to the use of different probes. If complexes B and C involve proteins which are not related to Spl, these proteins could either be previously identified GC box binding proteins or novel proteins.

In addition to Spl several other proteins which bind to GC box sequences have been identified. LSF (late SV40 factor) interacts with the SV40 GC boxes 2 and 3 to activate transcription (Huang et al., 1990). Binding requires two directly repeated GC motifs with a centre to centre spacing of 10 bp. LSF is therefore unlikely to interact with the PCMV LTR, which has only a single GC box motif. ETF interacts with GC-rich sequences to stimulate the transcription of the epidermal growth factor receptor gene, which does not have a TATA box sequence in the promoter (Kageyama et al., 1989). However, ETF does not activate transcription from promoters which do have a TATA box, such as the PCMV LTR promoter.

GCF (GC factor) is a human transcription factor which was cloned by Kageyama and Pastan (1989). The cDNA encodes a 91kd protein which binds to GC-rich sequences via a basic N-terminal domain. This factor represses transcription both in vivo and in vitro from the promoters
of several different genes. GCF will interact with a consensus Spl binding site although it has higher affinity for a variant sequence. MTF-1 is a zinc-inducible factor which interacts with the metal responsive element (MRE) of the mouse metallothionein-I gene (Westin and Schaffner, 1988). The MRE motif has homology with the consensus Spl binding site; Spl protein can interact with the metallothionein-I gene MRE. However, MTF-1 is distinct from Spl as higher zinc concentrations are required for MTF-1 binding, and methylation interference analysis has shown that the two proteins have very similar but non-identical binding sites. In the gel retardation experiments described in this study no zinc ions were added; if complexes B or C involve MTF-1 it must be postulated that the protein chelates sufficient zinc prior to extraction to remain functional. Westin and Schaffner (1988) have suggested that MTF-1 is related to Spl.

7.3 FURTHER CONSIDERATION OF THE RESULTS OF TRANSIENT TRANSFECTION EXPERIMENTS

Following the in vitro characterization of the proteins interacting with the -345 and -166 sites of the Mo-MSV LTR and its mutant derivatives, it was decided to assess the functional significance of these interactions by transient transfection analysis. The effect of the point mutations at these sites on efficiency of transcription were assessed by introducing the mutations separately and together into a construct in which the Mo-MSV LTR directs transcription of the cat gene.

The creation of a consensus Spl binding site, by introduction of the -166 point mutation to the LTR, increased transcription by 6-fold in F9 EC cells. This result implies that the interaction of F9 proteins with this site activated transcription. The results of Darrow et al. (1990) support this conclusion; they found that mutation of the consensus Spl binding sites in the t-PA promoter decreases its in vitro transcription levels in F9 EC cell extracts. It is most likely that the activation of Mo-MSV transcription on addition of an Spl binding site is a function of the murine Spl-like
protein. However, it cannot be ruled out that the proteins involved in complexes B and C, which have not been positively identified, are responsible for increased transcription. The activation of transcription on creation of the consensus Spl binding site strongly suggests that neither of these complexes involves GCF (Kagayama and Pastan, 1989) as this protein has been defined as a repressor of transcription.

The -166 point mutation facilitated a 2-fold increase in transcription efficiency in differentiated cells. This relatively low increase suggests that other activators of transcription had already elevated the efficiency of transcription closer to the point at which general transcription factors become limiting. The -166 point mutation did not allow increased transcription in PCC4 EC cells. This result was unexpected, as extracts of PCC4 and F9 cells contained similar concentrations of the proteins which interact with the -166 sequence.

Pugh and Tjian (1990) have recently presented evidence for the requirement of a non DNA-binding coactivator protein for Spl function. The coactivator was found to be present in partially purified HeLa cell TFIID preparations; further purification of TFIID prevented Spl activity. The coactivator is believed to act as an adaptor molecule, allowing Spl to interact with the general transcription machinery. One possible explanation of the inactivity of the -166 site binding proteins of PCC4 cells, is that this cell type is deficient in the required coactivator. A second explanation is that a PCC4-specific repressor of transcription successfully prevents transcriptional activators from functioning. Such an over-riding "shut-down" of transcription is not the favoured hypothesis as the results of Weiher et al. (1987) showed that the MPSV enhancer sequence is capable of directing efficient transcription in PCC4 cells. This requires addition of the B2 mutation (Barklis et al., 1986) within the negative regulatory element (NRE) in the 5' untranslated leader sequence. The NRE binding protein occurs in both F9 and PCC4 cells.

Deletion of the Mo-MSV enhancer sequences was found to have no effect on transcription from the LTR in F9 EC cells in this study.
This result conflicts with the data of Gorman et al. (1985a; 1985b), which shows that the promoter of Mo-MSV alone has a low, but significant, activity in F9 EC cells and that this activity is further reduced if the enhancer sequence is also present. This result was important as it led directly to the suggestion that EC cells contain a repressor of retrovirus transcription. However, other groups have also reported that the enhancer is functionless in EC cells (Linney et al., 1984).

The efficiency of transfection (0.5 to 5%) achieved in the experiments reported in this study is at least an order of magnitude lower than that reported by Gorman et al. (1985a; 1985b). The reason for this large difference in transfection efficiency is unclear as the transfection method, and cell line used, are apparently identical. However, as this is the only obvious difference between the experiments, it seems likely to account for the different results obtained. A model which can explain such an effect is proposed below.

If repressor protein(s) is present at a low concentration then transfection of high concentrations of DNA may titrate out all available repressor. The remaining DNA molecules, not bound by repressor protein, will have a similar transcriptional activity to sequences unable to bind the repressor (e.g. constructs in which the repressor binding site(s) has been deleted), hence differences in transcriptional activity of constructs in which repressor binding sites are present or absent will be masked. When transfection efficiency is low, the available DNA is able to enter only a small number of cells, hence the DNA concentration per cell is high allowing the scenario outlined above to occur. However, when transfection efficiency is high, the available DNA can enter a high number of cells and DNA concentrations per cell remain low, revealing the effects of repressor proteins present at low concentration. In addition, high efficiency of transfection allows significant levels of CAT activity to be detected even when very low absolute concentrations of DNA are transfected, thus allowing the DNA concentration per cell to be reduced still further.

The -345 point mutation was not found to affect transcription from the LTR in F9 or PCC4 EC cells. However, Tsukiyama et al. (1990)
found that the presence of eight tandem copies of the EC-R binding site, placed at the -345 site in the Mo-MuLV LTR, reduced transcription 3-fold on transient transfection of PCC4 cells. As the effect of eight copies of the binding site is so small it is not surprising that a single copy of the site should have an undetectable significance, as found in this study. Nevertheless, the fact that any effect at all was observed does suggest a functional role for the EC-R protein.

The model outlined above, which suggests that the inability to observe repression of transcription is a function of low transfection efficiency, can explain the inability to observe a role for EC-R. Thus, it cannot be concluded from the experiments presented here that EC-R is not a repressor of transcription. The transient transfection system is also not ideal for the study of retroviral transcription, as the DNA remains autosomal rather than integrating into the chromosome. When within the chromosome proviral DNA can be packaged into chromatin and such tertiary structure may be important to EC-R function.

Co-transfection experiments confirmed the hypothesis that the Ad5 E1A gene products down-regulate transcription from the Mo-MSV enhancer. The question of which product, the 243 or the 289 amino acid E1A protein, is responsible for this down-regulation of retroviral transcription has not yet been addressed. Velcich and Ziff (1985) reported similar results using the papovavirus SV40. In this case both the 243 and the 289 amino acid forms of E1A were capable of mediating down-regulation, although the mechanisms were not investigated.

The kinetics of down-regulation of transcription from the Mo-MSV LTR were not extensively studied; however, it is apparent that far higher concentrations of E1A expression plasmid were required to achieve a similar effect to that reported for SV40. It is impossible to make any direct comparisons between the experiments of Velcich and Ziff (1985) and the results reported in this study, as different cell lines and expression vectors were used. The apparent disparity in the results does not therefore imply that down-regulation of SV40 and retroviral transcription are achieved by different mechanisms.
7.4. A MODEL TO EXPLAIN TRANSCRIPTIONAL ACTIVITY OF MUTANT RETROVIRUSES IN EC CELLS

Considered as a whole, the results of these and previously published studies lead to the formulation of the following general model concerning retrovirus infection of early embryonic cells. For a productive infection to occur transcription must reach a currently undefined threshold level. In order to achieve this level in embryonic cells the wild-type LTR sequences must be mutated in a fashion that increases the binding efficiency of activators of transcription, decreases the binding efficiency of repressors of transcription or causes a combination of the two. The mutations capable of achieving these aims will vary between different cell types because different cells have a different complement of available factors.

Mutations in the negative regulatory element (NRE), which lies in the 5' untranslated leader sequence, decrease the binding efficiency of an EC cell-specific repressor of transcription (Loh et al., 1990). These mutations include B2 (Barklis et al., 1986) which allows efficient transcription in the absence of any other mutations. MPSV also has mutations in the vicinity of the NRE, which are required in addition to the mutations 5' of the transcriptional start for efficient transcription (Weiher et al., 1987). Presumably the MPSV mutations in the NRE decrease but do not abolish repressor binding. In support of this hypothesis Weiher et al. (1987) showed that addition of the B2 mutation to MPSV increases the efficiency with which it is transcribed in EC cells.

Creation of an Spl binding site in MPSV by a single point mutation allows increased efficiency of transcription in F9 EC cells. However, there is no reason to believe that mutations which form binding sites for other transcriptional activators present in EC cells would not have a similar positive effect. It is probably merely fortuitous that a site with partial identity to a consensus Spl binding site was already present in the Mo-MSV LTR. EC-R is potentially a second repressor of transcription. However, the effect of disrupting the EC-
R binding site upon viral transcription (as opposed to CAT construct transcription) remains to be seen.

The NRE represses transcription to such a level that even the presence of a functional Sp1 binding site in the promoter cannot over-ride its effect and allow a productive infection to occur. This was demonstrated by Weiher et al. (1987) in an experiment in which the U3 region of Mo-MuLV (containing the enhancer and promoter sequences) was replaced by that of MPSV. The mutant construct remained incapable of productively infecting EC cells. This result does not prove that the NRE is always dominant, as addition of further "positive" sequences in the U3 sequence might allow the required threshold level of transcription to be reached and a productive infection to occur.

7.5. PROSPECTS FOR FUTURE STUDY

The experiments reported in this study have yielded the information required for an updated version of the highly informative viral infection experiments of Weiher et al. (1987). The approach of Weiher et al. (1987) produced results which reflected the true regulation of transcription in a provirus. Approaches using transient transfection of CAT constructs have revealed effects of the promoter and enhancer sequences, but have not investigated the effects of the control sequences located in the untranslated 5' leader.

By performing infection experiments using whole retroviruses, into which the mutations considered in this study may be introduced, a number of potential problems would be eliminated.

1) The low efficiency of introduction of DNA to EC cells by transfection would be circumvented by use of infection.

2) The problem of too high a concentration of DNA entering individual cells could also be eliminated, by use of a low titre of virus to ensure an average of one viral genome per cell.

3) Potential cooperation between control sequences 5' and 3' of the transcriptional start could be investigated by looking at the levels of transcription in the presence and absence of the B2 mutation (Barklis et al., 1986).
4) Use of a provirus allows the investigation of transcriptional regulation in the presence of chromatin structure.

In order to perform such experiments the plasmid MFGLacZ (a kind gift from R. Mulligan, Whitehead Institute, Cambridge, USA), which includes wild-type Mo-MuLV LTR sequences, primer binding sites and packaging signals plus LacZ as a marker gene, was used as a starting point to make a variety of different mutant viruses. If the 3' LTR is replaced with a mutant version, the 5' LTR of the provirus will obtain the same sequence during reverse transcription. The NheI to XbaI fragment from -409 to -152, with respect to the transcriptional start, of the 3' LTR of MFGLacZ was therefore replaced with the corresponding sequences from pMSVcat, pMSV(-166)cat, pMSV(-345)cat or pMSV(-166,-345)cat. The B2 mutation was also introduced into each construct by site-directed mutagenesis.

Unfortunately, due to time constraints, analysis of these constructs has not yet been carried out. Each construct will be transfected into a packaging cell line, such as the Psi2 line (Mann et al., 1983), which provides virion proteins enabling the mutant viral RNAs to be packaged. These mutant viruses could be produced transiently or stable producer lines isolated. Each virus will then be titred on a permissive cell type before being used to infect F9 and PCC4 EC cells. Efficiency of transcription may then be determined by assaying for β-galactosidase activity.

This type of experiment should be particularly revealing as it enables the effect of each point mutation to be analyzed in the presence or absence of each other point mutation. Thus, interactions between the different control elements may be revealed. The results of this experiment would also be useful in the design of new viral vectors. Retroviral infection is an extremely useful method for transferring DNA into cells, however it has been difficult to use this method with ES cells due to inefficiency of transcription. Positive identification of an efficiently transcribed LTR sequence would enable a more successful vector to be constructed.

It has not yet been shown that the EC-R protein is a repressor of transcription. However, assuming that the viral infection experiments confirm this hypothesis many interesting experiments could be
performed. The results of this study have shown that EC-R can be purified using oligonucleotide affinity chromatography. Pure protein could be used to clone the EC-R gene by raising antibodies to screen a c-DNA expression library. Alternatively, peptide sequencing could be used to predict DNA sequence, from which oligonucleotides could be designed to screen a c-DNA library by hybridization. Once cloned the expression pattern of the gene during early development could be analyzed and its potential role in developmental control investigated.

Finally, it has been demonstrated in this study that E1A gene products down-regulate the transcription of retroviruses. Thus, retroviruses can provide a further experimental system, in addition to the papovaviruses and cellular genes, in which to study the mechanism of E1A-mediated repression.

In conclusion, the original aims of the project have been at least partially fulfilled. The approach taken to investigate the control of transcription in pluripotential embryonic stem cells, using a viral promoter previously shown to be differentiation-regulated, has proved an effective one. It has been shown that a previously identified activator of transcription is both present and functional in embryonal carcinoma stem cells. In addition a potential repressor of stem cell transcription has been identified. Some light has been shed upon the block to retrovirus transcription in early embryonic stem cells, and a route by which this phenomenon may be further investigated has been revealed.
ACKNOWLEDGEMENTS

I would like to thank my supervisor Peter Rigby and co-supervisor David Latchman for their advice, encouragement and enthusiasm during the course of this work.

I am indebted to the following for their generous gifts: David Stott for the HPRTcat plasmid, Davor Solter for the antibody raised against stage specific embryonic antigen, Steve Jackson for anti-Spl antibody and purified HeLa cell protein, Nic Jones for the E1A expression plasmid and anti-E1A antibody, and Richard Mulligan for the MFGlacZ construct.

I would also like to thank Panayotis Tassios and Eugenia Spanopoulou for sharing their unpublished observations, Rob Hayes for assistance with FPLC analysis, and Jack Price and Rob Moore for advice on "retrovirology". Thanks are also due to the various technical support groups at NIMR and in particular to Chris Pratchett for plasmid preparations, to Mike Furness and Diane Chase for oligonucleotide synthesis, to the photography and graphics departments, and to Dave Cox and Paula Timmons for maintaining the smooth running of the laboratory.

I am very grateful to David Stott, Ted Murray, Bob White, Janet Partridge, Sui-Pok Yee, Ruedi Alleman, Panayotis Tassios, Tina Dalianis and Nick La Thangue for advice on techniques and useful discussions.


273


276


279


Piette, J. and Yaniv, M. (1987). Two different factors bind to the α-domain of the polyoma virus enhancer, one of which also interacts with the SV40 and c-fos enhancers. *EMBO J.* 6, 1331-1337.


290


