DEVELOPMENTAL EXPRESSION OF A NOVEL MURINE GENE ISOLATED FROM EMBRYONAL CARCINOMA CELLS.

A thesis presented by

Alastair D Reith

for the degree of Doctor of Philosophy

at

Imperial College of Science and Technology,

University of London

(April 1988)

ABSTRACT.

Differential screening of a cDNA library prepared from the pluripotential murine embryonal carcinoma cell line PCC3 had led to the isolation of a number of clones representing genes the expression of which decreased on *in vitro* differentiation of this cell line. The expression pattern of a number of these genes was determined in order to establish which were suitable for the analysis of gene function and regulation during mouse embryogenesis.

One gene, GR1, was found to encode a 6.8kb mRNA abundant in PCC3 cells but detectable only in trace amounts following retinoic acid-induced differentiation. Northern blot analyses established that GR1 RNA was present in mid-gestation mouse embryos but not in a variety of adult tissues, with the exception of trace amounts in adult brain and testis.

The nucleotide sequence of 2.67kb of the GR1 transcription unit was determined. This sequence was found to include 247 amino acids of the carboxy terminal region of the putative GR1 polypeptide. Both nucleotide and amino acid sequences were novel. The GR1 locus was recovered from genomic libraries only when an unamplified bacteriophage lambda genomic library was propagated on a methylcytosine restriction deficient ($McrB^-$) host *E.coli* strain. Use of the BxD series of recombinant inbred mice established linkage between the GR1 locus and the *met* proto-oncogene on mouse chromosome 6.

RNA *in situ* hybridisation analyses revealed GR1 expression at

a number of discrete sites during mouse embryogenesis. Whilst expressed in presomitic mesenchyme, GR1 RNA localised to the sclerotome of somites and, subsequently, to the anterior sclerotome half in the 12.5 days gestation embryo. Expression was also seen in the cartilaginous tissues of the embryonic axial skeleton and in developing connective tissues and craniofacial structures. High levels of GR1 RNA were detected in the developing anterior lobe of the pituitary gland and a number of extraembryonic cell types.

The sites of GR1 expression, along with analysis of the partial amino acid sequence of the putative GR1 polypeptide, are consistent with the notion that this gene encodes a novel component of some extracellular matrices.

Acknowledgements.

I would like to thank Peter Rigby for his advice and enthusiasm in the supervision of this project and Ashley Cooper for discussions on many aspects of this work.

I am indebted to Bill Brammar, Frank Grosveld, Bernhard Herrmann, Ian Jackson, Peter Little, Robin Lovell-Badge, Kam Man Hui and Will Whitfield for gifts of various reagents and Nick Clark, Peter Gillett, Kevin O'Hare, Brigid Hogan, David Leach, Ted Murray, Mike Snow, Ben Taylor, Willie Taylor, and David Wilkinson for useful discussions on various techniques and analyses.

I am also grateful to Joe Brock, Neil Papworth and all at the photography section, NIMR, for assistance with art work, and Lars Ahrlund-Richter and Debbie Duthie for assistance with word processing.

Most importantly, I would like to thank my parents, Gillian and Seamus McKenna, and Christine Ellis for their love and unfailing support, without which none of this work would have been possible.

To my parents, Ella and Tom Reith, for their love and guidance.

CONTENTS.

Chapter 1. General Introduction.

Section 1.1. Normal development of the mouse.

Section 1.2. In vitro models for the early stages of mouse embryogenesis.

1.2.1. Embryo-derived stem cells.

1.2.2. In vitro differentiation of EC and ES cells.

Section 1.3. Gene regulation in murine embryonal carcinoma cells.

1.3.1. Genes regulated in EC cells.

1.3.2. Transcriptional regulation of gene expression in murine EC cells.

Section 1.4. Analysis of gene function and regulation during mouse embryogenesis.

Chapter 2. Materials and Methods.

Section 2.1. Bacterial strains.

Section 2.2. Bacteriophage strains.

Section 2.3. Mammalian cell culture.

Section 2.4. Animals.

Section 2.5. Large scale preparation of bacteriophage lambda.

Section 2.6. Large scale preparation of cosmid and plasmid DNA.

Section 2.7. Small scale preparation of cosmid and plasmid DNA.

Section 2.8. Preparation of mammalian genomic DNA.

Section 2.9. Preparation of total cellular RNA.

Section 2.10. Preparation of total cytoplasmic RNA.

Section 2.11. Selection of polyadenylated RNA on oligodT-cellulose.

Section 2.12. Preparation of crude protein extracts.

Section 2.13. Quantitation of nucleic acids.

Section 2.14. Restriction enzyme digestions.

Section 2.15. Electrophoretic techniques.

Section 2.16. Molecular cloning techniques.

Section 2.17. Radiolabelling of nucleic acids.

Section 2.18. Filter preparation for nucleic acid hybridisations.

Section 2.19. Embryo and tissue preparation for RNA in situ hybridisation.

Section 2.20. Hybridisation techniques.

Section 2.21. Autoradiography.

Section 2.22. Sequencing of DNA.

Section 2.23. Transfer of proteins from SDS-PAGE to nitrocellulose membranes.

Section 2.24. Raising of polyclonal antisera.

Section 2.25. Western blot analysis.

Section 2.26. Immunocytochemical staining.

Chapter 3. Selection of cDNA clones suitable for studies of gene regulation in embryonic stem cells.

Section 3.1. Introduction.

- 3.1.1. Strategies for the isolation of differentially regulated genes.
- 3.1.2. Criteria for genes suitable for transcriptional analysis.

Section 3.2. Results.

- 3.2.1. Expression of 1.6U.
- 3.2.2. Transcriptional orientation of p1.6.
- 3.2.3. DNA sequence analysis of p1.6.
- 3.2.4. Expression of λ22 in murine embryonal carcinoma cells.
- 3.2.5. DNA sequence analysis of p2265.
- 3.2.6. Expression of GR4 in murine embryonal carcinoma cells.
- 3.2.7. Expression of GR1 in murine embryonal carcinoma cells.
- 3.2.8. Kinetics of GR1 regulation on *in vitro* differentiation of PCC3 cells.
- 3.2.9. GR1 expression in embryonic stem cells and mouse embryogenesis.
- 3.2.10. GR1 expression in embryonic, neonatal and adult mouse tissues.

Section 3.3. Discussion.

- 3.3.1. Characterisation of 1.6U.

3.3.2. Characterisation of λ22.

3.3.3. Expression pattern of GR4.

3.3.4. Regulated expression of GR1.

Chapter 4. Cloning and preliminary characterisation of the GR1 locus.

Section 4.1. Introduction.

4.1.1. Cloning eukaryotic genomes in Escherichia coli.

4.1.2. Instability of cloned genomic sequences.

4.1.3. Factors influencing the stability of cloned genomic sequences.

Section 4.2. Results.

4.2.1. Restriction mapping of GR1 locus.

4.2.2. A moderately repetitive element within pGR165.

4.2.3. Screening amplified libraries for GR1 genomic clones.

4.2.4. Analysis of pGR165 homologous cosmid clones.

4.2.5. Construction of mouse genomic library in bacteriophage lambda.

4.2.6. Isolation of pGR165 homologous genomic clones.

4.2.7. Isolation of 1.6U homologous genomic clones.

Section 4.3. Discussion.

4.3.1. Characterisation of the GR1 locus.

4.3.2. Cloning the GR1 locus.

4.3.3. GR1 genomic clones.

4.3.4. 1.6U genomic clones.

Chapter 5. RNA in situ hybridisation and fusion protein analyses of GR1.

Section 5.1. Introduction.

5.1.1. Analysis of GR1 expression during mouse development.

5.1.2. RNA in situ hybridisation analyses.

5.1.3. Raising antisera against putative GR1 proteins.

Section 5.2. Results.

5.2.1. Transcriptional orientation of pGR165.

5.2.2. Production of β-galactosidase-GR1 fusion proteins in E.coli.

5.2.3. Polyclonal rabbit antisera against the 291-12 fusion protein.

5.2.4. RNA in situ hybridisation analyses.

(i) GR1 expression in extraembryonic tissues.

(ii) GR1 expression during gastrulation.

(iii) GR1 expression in sclerotome and derived structures.

(iv) GR1 expression in other embryonic skeletal tissues.

(v) GR1 expression in lung and gut mesenchyme and connective tissues.

(vi) GR1 expression in developing neural tissues.

(vii) Localisation of GR1 RNA in embryonic and adult testis.

Section 5.3. Discussion.

5.3.1. GR1 fusion protein studies.

5.3.2. Expression of GR1 during mouse development.

(i) Relating GR1 expression in the early embryo to embryo-derived stem cell lines.

(ii) Expression during somitogenesis and bone development.

(iii) Expression in connective tissues and other mesenchyme derivatives.

(iv) GR1 expression in craniofacial structures.

5.3.3. A role for GR1 in mouse development?

Chapter 6. Chromosome mapping and DNA sequence analyses of GR1.

Section 6.1. Introduction.

6.1.1. Determination of nucleic acid sequences.

6.1.2. Analysis of DNA and amino acid sequences.

6.1.3. Linkage map of the mouse.

Section 6.2. Results.

6.2.1. GR1 cDNA clones available for DNA sequence determination.

6.2.2. Sequencing strategy for pGR165

6.2.3. Determination of partial DNA sequence of GR1 genomic clone N28.

6.2.4. Analysis of partial DNA sequence of the GR1 transcription unit.

6.2.5. Analysis of partial amino acid sequence of the putative GR1 polypeptide.

6.2.6. Mapping the GR1 locus in BxD recombinant inbred mice.

Section 6.3. Discussion.

6.3.1. Nucleotide sequence analyses of GR1 transcription unit.

6.3.2. Analyses of putative GR1 polypeptide.

6.3.3. Chromosome mapping of GR1 locus.

Chapter 7. General discussion and future prospects.

Section 7.1. Regulation of GR1 expression in EC cells and during mouse development.

7.1.1. Mechanism of transcriptional regulation of GR1 expression.

7.1.2. Transgenesis as a route to the identification of GR1 *cis*-regulatory elements.

Section 7.2. Function of GR1 in mouse development.

7.2.1. Expression pattern and sequence analyses of GR1.

7.2.2. Antisera against GR1 protein(s).

7.2.3. Poly-proline encoding repetitive element in the putative GR1 polypeptide.

7.2.4. Analysis of GR1 function in transgenic mice.

7.2.5. Diversification of analyses into other experimental systems.

LIST OF FIGURES.

Fig. 1.1. Early development of the mouse.

Fig. 1.2. Origin of murine embryonal carcinoma cells.

Fig. 1.3. *In vitro* differentiation of EC and ES cells.

Fig. 3.1. Restriction map of p1.6.

Fig. 3.2. Northern blot analysis of 1.6U expression in EC cells.

Fig. 3.3. Transcriptional orientation of p1.6.

Fig. 3.4. Partial nucleotide sequence of p1.6.

Fig. 3.5. Restriction map of λ22.

Fig. 3.6. Northern blot analysis of λ22 expression in EC cells.

Fig. 3.7. Partial nucleotide sequence of p2265.

Fig. 3.8. Northern blot analysis of GR4 expression in EC cells.

Fig. 3.9. Restriction map of pGR165.

Fig. 3.10. Northern blot analysis of GR1 expression in EC cells.

Fig. 3.11. Kinetics of GR1 regulation on differentiation of PCC3 cells.

Fig. 3.12. Northern blot analysis of GR1 expression in ES cells and embryogenesis.

Fig. 3.13. Northern blot analysis of GR1 expression in embryonic, newborn and adult tissues.

Fig. 3.14. Northern blot analysis of GR1 expression in adult mouse tissues.

Fig. 4.1. Potential problems in cloning mammalian genomic loci.

Fig. 4.2. cDNA probes used for mapping GR1 genomic locus.

Fig. 4.3. Restriction map of GR1 locus.

Fig. 4.4. Southern blot analysis of a repetitive element in pGR165.

Fig. 4.5. Screening amplified genomic libraries for GR1 genomic clones.

Fig. 4.6. Restriction maps of CAD18, CAD20 and CAD22 genomic clones.

Fig. 4.7. Efficient ligation of genomic and bacteriophage vector restriction fragments.

Fig. 4.8. Construction of bacteriophage lambda library from PCC3 genomic DNA.

Fig. 4.9. Southern blot analysis of N28 genomic clone.

Fig. 4.10. Restriction map of N28 genomic clone.

Fig. 4.11. Southern blot analysis of N20.1 and N20.3 genomic clones.

Fig. 4.12. Southern blot analysis of 1.6U genomic clones.

Fig. 4.13. Titre of 1.6U genomic clones on different E.coli strains.

Fig. 5.1. Construction of lacZ-GR1 fusion protein encoding plasmids.

Fig. 5.2. Transcriptional orientation and fusion protein analysis of pGR165.

Fig. 5.3 GR1 expression in 7.5 days gestation mouse conceptus.

Fig. 5.4. GR1 expression in mid-gestation placenta.

Fig. 5.5. Localisation of GR1 RNA in 8.5 days gestation mouse embryo.

Fig. 5.6. GR1 expression in 9.5 days gestation mouse embryos.

Fig. 5.7. Origin of vertebral segments.

Fig. 5.8. Localisation of GR1 expression in sclerotome and derivatives.

Fig. 5.9. GR1 expression in developing limbs.

Fig. 5.10. GR1 expression in embryonic ribs.

Fig. 5.11. GR1 expression in gut and lung mesenchyme.

Fig. 5.12. GR1 expression in developing craniofacial bones and connective tissue.

Fig. 5.13. Localisation of GR1 RNA in fetal brain and spinal cord.

Fig. 5.14. GR1 expression in developing pituitary gland.

Fig. 5.15. Localisation of GR1 RNA in embryonic and adult mouse testis.

Fig. 6.1. Sequencing strategy for cDNA clone pGR165.

Fig. 6.2. Nucleotide sequence of cDNA clone pGR165.

Fig. 6.3. Sequencing strategy for genomic subclone pBS28A.

Fig. 6.4. Partial nucleotide sequence of the GR1 transcription unit and putative GR1 polypeptide.

Fig. 6.5. Amino acid composition of putative GR1 polypeptide.

Fig. 6.6. Predicted secondary structure of putative GR1 polypeptide.

Fig. 6.7. Restriction fragment length polymorphisms at GR1 locus.

Fig. 6.8. Strain distribution pattern of the GR1 locus in BxD recombinant inbred mice.

Fig. 6.9. Potential termination site of pGR165 and open reading frames of pUR-12 fusion proteins.

Fig. 6.10. Location of GR1 locus on mouse chromosome 6.

LIST OF TABLES.

Table 2.1. Bacterial strains used for growth of plasmids and bacteriophage.

Table 2.2. Media used for growth of E.coli.

Table 2.3 Bacteriophage and plasmids.

Table 2.4. Mammalian cell culture.

Table 2.5. Conditions for restriction enzyme digestion of DNA.

Table 2.6. Electrophoretic buffers.

Table 2.7. Reagents for dideoxy sequencing using Klenow fragment of E.coli DNA polymerase I.

Table 2.8. Reagents for dideoxy sequencing using modified T7 DNA polymerase.

Chapter 1. General Introduction.

Understanding the processes responsible for the orderly morphogenesis of the fertilised egg into a complex and highly differentiated multicellular organism is one of the central problems of molecular biology. The fertilised egg contains the entire complement of genetic information required to direct orderly development, by means of appropriate spatial and temporal patterns of gene expression. A given gene product may be required at a particular stage in development and be restricted to certain lineages or cell types.

Understanding of the mechanisms of mammalian development will provide insights into the question of how the genetic information within each cell is utilised to form organisms as complex as ourselves. In addition, knowledge of how genes regulate growth and differentiation of the mammalian embryo is essential to facilitate an understanding of human congenital defects and abnormal development and will also provide opportunities for the genetic improvement of many economically important mammalian species and diversification of farming practice.

The mammalian embryo is uniquely adapted for development within the protective and nutritive environment of the mother. Whilst this limits the usefulness of other vertebrate species in the analysis of mammalian development, it makes the mouse embryo less suitable for experimental manipulation than Xenopus or the chick, two species used extensively in the study of vertebrate development.

Of the mammals, the mouse is the only species for which

genetics has advanced to such a degree as to allow the study of development, a reflection of the small size and relatively short generation time of this species. The recent development of super-ovulation, embryo culture and germline manipulation techniques have overcome, to some extent, the problems of experimental manipulation of the mouse embryo. In conjunction with the use of *in vitro* cell cultures representative of a variety of mammalian cell types, rapid advances in knowledge of gene function and regulation during mammalian development have been made.

It is the combination of different gene products which determine, to a large extent, the properties of a cell and distinguish it from other cell types. Restriction of a given gene product to specific spatial and/or temporal locations within the embryo could be the result of regulation at one or more of the steps involved in the conversion of the genetic information within nucleotide sequences to functional proteins. In addition to regulating the occurrence of a given gene product, the activity of that product is likely to be influenced by the absence or presence of other gene products within a given cell.

One of these levels of regulation of gene expression is that of differential rates of transcription in the nucleus. Transcriptional regulation of gene expression is amenable to experimental analyses in a number of ways. Such experimental studies have three basic requirements:

(i) The availability of cloned genes which show transcriptional regulation in the system of interest.

(ii) Efficient methods for the introduction of mutated gene constructs into the system of interest, in order to define the cis-acting regulatory elements of the gene.

(iii) In vitro assay systems, which faithfully mimic the regulation seen in vivo, to facilitate the identification and purification of trans-acting factors responsible for the differential transcription of the gene under investigation.

In addition to investigating the mechanism by which differential expression of the gene is achieved, it is also pertinent to question why that gene is regulated within the developing embryo. Such a question centres on the function of the gene product which, in turn, is related to the properties of the protein encoded by the gene under investigation.

Relevant properties include cellular and tissue localisation, and the structure of the protein, it being a central tenet of molecular biology that structure is related to function. As with transcriptional regulation studies, the ability to introduce mutated gene constructs into the experimental system is an important feature of any functional analysis. The consequences for the developing embryo of deregulated expression of the gene, and/or normal expression of mutant gene products, can provide valuable information concerning the role played by that gene during development.

Whilst the mouse provides the best available model for mammalian development, there are a number of practical

limitations to the use of the mouse embryo for experimental analyses. As a consequence, a number of compromises must be made in order to approach questions concerned with transcriptional mechanisms of gene regulation, and gene function, during the early stages of mouse embryogenesis.
The information and reagents obtained from such work may subsequently prove useful in alleviating some of the experimental limitations currently imposed on the molecular analysis of mammalian development. It is with such current limitations, compromises and development of novel techniques that the following sections are concerned.

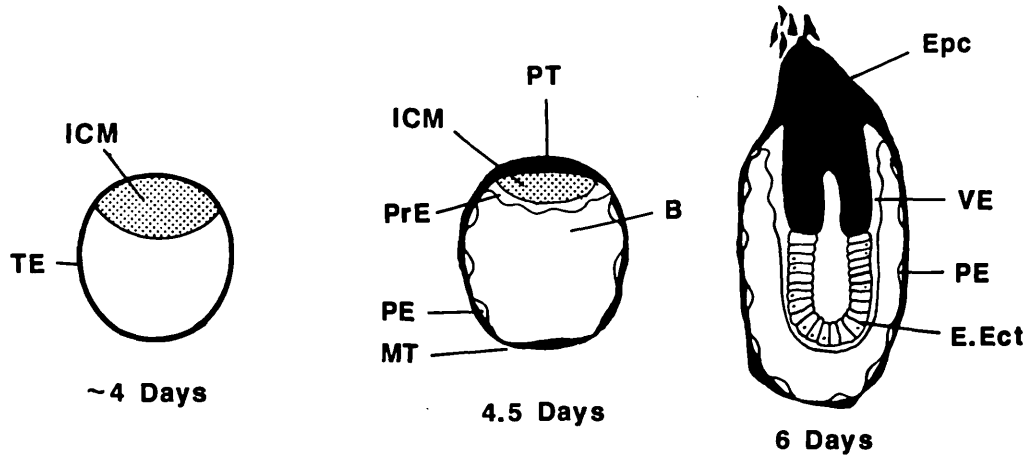
Section 1.1. Normal mouse development.

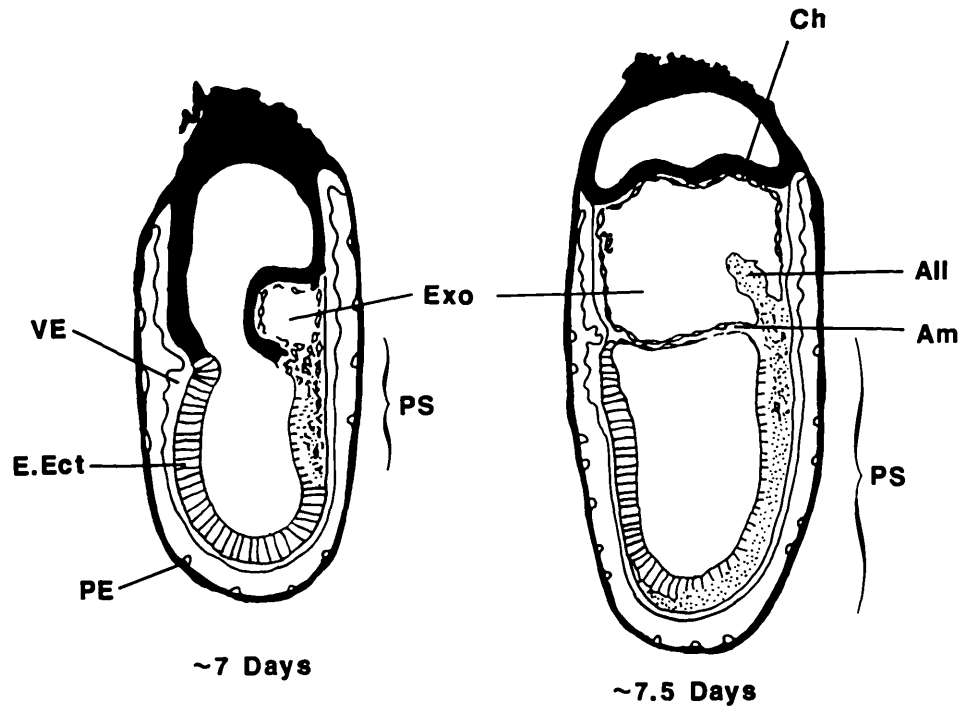
Some of the practical limitations to the use of the mouse as a model system for the study of gene function and regulation during the early stages of mammalian embryogenesis are apparent on consideration of the normal development of the mouse. An outline of the early stages of mouse development are given in figure 1.1. With regard to the study of the mechanisms of gene regulation, clearly it is the pluripotential cells within the embryo which are of major interest. From this population of cells the wide variety of differentiated cell types within the later embryo arises.
Following fertilisation, the zygote undergoes a series of symmetrical divisions to generate a morula consisting of

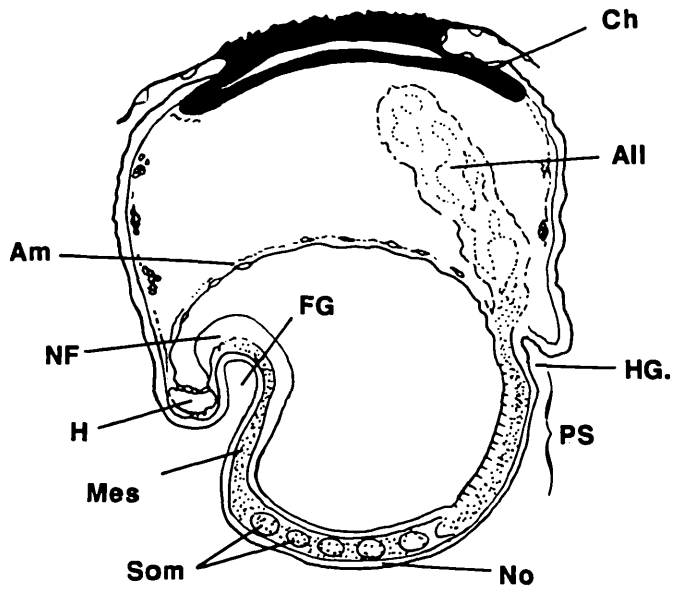
Figure 1.1. Early stages of mouse embryogenesis.

The development of the mouse embryo between approximately 4 days gestation and 8.5 days gestation are illustrated.

All= allantois, Am= amnion, B= blastocoel, Ch= chorion, E.Ect= embryonic ectoderm, Epc= ectoplacental cone, Exo= exocoelom, FG= foregut, H= heart, HG= hindgut, ICM= inner cell mass, Mes= mesenchyme, MT= mural trophectoderm, NF= neural fold, No= notochord, PE= parietal endoderm, PrE= primitive endoderm, PS= primitive streak, PT= polar trophectoderm, Som= somite, TE= trophectoderm, VE= visceral endoderm.


ICM
TE
~4 Days
PT
ICM
PrE
B
PE
MT
4.5 Days
Epc
VE
PE
E.Ect
6 Days


VE
Exo
PS
E.Ect
PE
~7 Days
Ch
All
Am
PS
~7.5 Days


Ch
All
Am
FG
NF
HG.
H
PS
Mes
Som
No
8.5 Days

individual blastomeres. Evidence from a number of experiments (reviewed by Hogan et al., 1986) has established that up to the 2 cell stage, development proceeds by utilisation of maternal mRNA stored within the unfertilised egg. At the mid 2 cell stage the zygotic genome is activated, and by the late 2 cell stage the bulk of the maternal mRNA has been degraded (Piko and Clegg, 1982).

Up to the 8 cell stage there is good evidence to suggest that individual blastomeres are equipotent, but not totipotent. Individual blastomeres will not, by themselves, form a complete mouse but will efficiently contribute to many tissues when used in chimera formation with genetically distinct blastomeres (Kelly, 1977). Beyond this stage, there is a gradual restriction in the developmental potential of individual cells within the embryo.

At around 2.5 days post coitum, the morula undergoes a process known as compaction in which individual blastomeres become polarised and extensive gap-junctional communication is established between the cells of the embryo (Johnson and Maro, 1986). Polarisation involves the establishment of distinct apical and basal membrane domains, as well as polarisation within the cytoplasm. As a consequence of this process, asymmetry is established which ultimately results in the first differentiation events within the embryo by the 32-64 cell stage, the formation of trophectoderm and inner cell mass (ICM) which together constitute the blastocyst and appear at around 3.5-4 days gestation (fig. 1.1). Within the 64 cell

stage blastocyst, approximately 20 cells constitute the inner cell mass situated at one end of a fluid-filled blastocoel cavity formed by the surrounding trophectoderm cells (reviewed by Hogan et al., 1986).

Up to this stage, development has proceeded within a protective sheath, the zona pellucida, in the oviduct. The blastocyst now "hatches" from the zona and implants into the uterine wall.

Trophectoderm cells not in contact with the ICM, the mural trophectoderm, synthesise a trypsin-like enzyme, strypsin (Wassarman et al., 1984) which removes the zona pellucida. During implantation, at around 4-4.5 days gestation, the mural trophectoderm cells invade the uterine epithelium and penetrate the endometrium. The mural trophoblast cells subsequently undergo several rounds of endoreduplication, forming trophoblast giant cells which contribute to the parietal yolk sac at later stages of development.

Following implantation, some of the polar trophectoderm, the trophoblast in contact with the ICM, migrates, replacing the mural trophectoderm. These cells subsequently form trophoblast giant cells. Others remain diploid and proliferate rapidly, so pushing the ICM into the blastocoel cavity. These trophectoderm cells form the extraembryonic ectoderm of the egg cylinder stage, at 7 days gestation, and subsequently contribute to both chorionic ectoderm and ectoplacental cone portions of extraembryonic tissues (Gardner, 1983).

The 20 or so cells in the 64 cell stage blastocyst which

consitute the ICM also undergo a series of differentiation events. Primitive endoderm cells form on the blastocoel surface of the inner cell mass at about 4.0-4.5 days gestation, just prior to implantation (fig. 1.1). The primitive endoderm cells subsequently differentiate to two cell types. The parietal endoderm cells lie on the blastocoel surface of the trophectoderm whilst visceral endoderm cells lie on the blastocoel surface of the remaining ICM, now known as the embryonic ectoderm, primitive ectoderm or epiblast. Parietal and visceral endoderm cell types constitute the extraembryonic endoderm and subsequently form the parietal yolk sac and visceral yolk sac, respectively, at later stages of development (Gardner, 1983). These tissues, along with trophectoderm derivatives and extraembryonic mesoderm and extraembryonic ectoderm-derived cells, are concerned with protection and nourishment of the fetus.

There is considerable evidence from injection chimera experiments to suggest that the primitive ectoderm cells give rise to the three germ layers of the embryo, as well as extraembryonic ectoderm and mesoderm, whilst primitive endoderm forms only extraembryonic components (Gardner and Rossant, 1979; Gardner, 1982; 1983).

Transplantation of embryos, up to 6 days gestation, to ectopic sites results in the formation of teratocarcinomas containing cells of all three germ layers plus pluripotential embryonal carcinoma (EC) cells (Stevens, 1970). However, teratocarcinomas cannot be established from 7.5 days gestation

embryos. Moreover, whilst 4.5 days gestation primitive ectoderm cells can contribute to all three germ layers in chimeric mice (Gardner and Rossant, 1979), cells within 7 day gestation egg cylinders are restricted in their developmental capacity (Snow, 1981).

By 6.5 days gestation, there are approximately 1000 cells in the primitive ectoderm (Snow, 1977). It is from these cells that all three germ layers arise by the process of gastrulation. A number of morphogenic movements occur during gastrulation by which the main body plan of the embryo is established.

Gastrulation in the mouse is very similar to that of the chick (Hogan et al., 1985) and much of the knowledge concerning morphogenic movements associated with gastrulation has come from scanning electron microscopy data from both species (Tam, 1981; Meier and Tam, 1982; Tam and Meier, 1982).

Mesoderm cells arise from the epithelial layer of primitive ectoderm and come to lie between primitive ectoderm and visceral endoderm cells. The mesoderm arises initially at the posterior end of the primitive ectoderm and then extends anteriorly, as the primitive streak, resulting in the appearance of both paraxial and lateral mesoderm. At the anterior end of the primitive streak arises Henson's node. Primitive ectoderm cells which migrate through Henson's node are laid down as mesodermal cells which subsequently form the notochord, separating the mesodermal sheet.

Subsequently, the neural plate arises as a thickening of

ectoderm overlying the mesodermal sheet and the paraxial mesoderm on either side of the notochord condenses into a series of paired somites along the longitudinal axis of the embryo. This series of events occurs over a number of days in the mouse and occurs in an anterior to posterior gradient within the embryo. Thus, at 8.5 days gestation, anterior somites are condensing whilst new mesodermal cells are arising from the primitive streak at the posterior end of the embryo.

The neural plate, overlying the notochord and paraxial mesoderm, folds up and eventually closes to form the neural tube. Closure of the neural tube occurs initially at the level of the 4th-5th somite at 8.5 days post coitum and then proceeds in both anterior and posterior directions, closure being complete at around 9 days gestation (Theiler, 1972).

Within the somites, differentiation occurs to generate three regions. The dermatome forms the connective tissue of the dermal layer, the myotome gives rise to the muscles of the body wall, and sclerotome differentiates to form the vertebrae and ribs. Limb buds arise from the lateral mesoderm plate and subsequently form limb cartilage and bone.

Thus, by approximately 10.5 days gestation the body plan of the embryo is established.

Section 1.2. *In vitro* models for the early stages of mouse embryogenesis.

From a consideration of the requirements of a molecular analysis of gene regulation during early mouse embryogenesis, and the actual process of mouse development, some of the problems with such analyses are apparent.

The small size, asynchrony of early development, and variety of cell types even within the blastocyst, precludes the direct use of early mouse embryos as a source of the large quantities of pure cell populations required for transcriptional studies. The variety of cell types within the embryo could prevent the isolation of clones the expression of which is regulated only in a proportion of the cells of an embryo between two stages of development. This could result in the isolation of only a very limited number of clones as was found in differential screening of cDNA prepared from 8 days and 10 days gestation mouse embryos (Wilkinson *et al.*, 1987a).

Thus, both for the isolation of regulated genes and analysis of transcriptional regulatory mechanisms operating during early mouse development, there is a need for large amounts of homogeneous populations of embryonic cells which can be manipulated *in vitro*.

1.2.1. Murine embryo-derived stem cells.

The requirement for large amounts of pure populations of embryonic cells is met by the use of *in vitro* cell culture systems representative of the pluripotential embryonic stem cells in the mouse blastocyst (section 1.1). The availablilty of such cell lines, along with methods of inducing their differentiation *in vitro*, provides a suitable model system for the differentiation events which occur within pluripotential embryonic stem cells in the developing embryo.

As mentioned earlier, the ICM/primitive ectoderm cells of 4.5-7.0 days gestation embryos are pluripotential and can contribute to all three germ layers in chimeric mice (section 1.1). Additional evidence for such pluripotential cell types in the early embryo came from the development of transplantable teratocarcinomas (Stevens, 1970).

Teratocarcinomas can arise spontaneously in 129/SV strain of mice or can be induced experimentally by grafting of primordial germ cells (Stevens, 1967) or early embryos (Stevens, 1970) to ectopic sites, usually beneath the kidney capsule or the testis. There is good evidence that transplantable teratocarcinomas induced by transplantation of early embryos to ectopic sites are derived from pluripotential embryonic cells (Stevens, 1970), and that it is the embryonic ectoderm which contains these cell types (Diwan and Stevens, 1976).

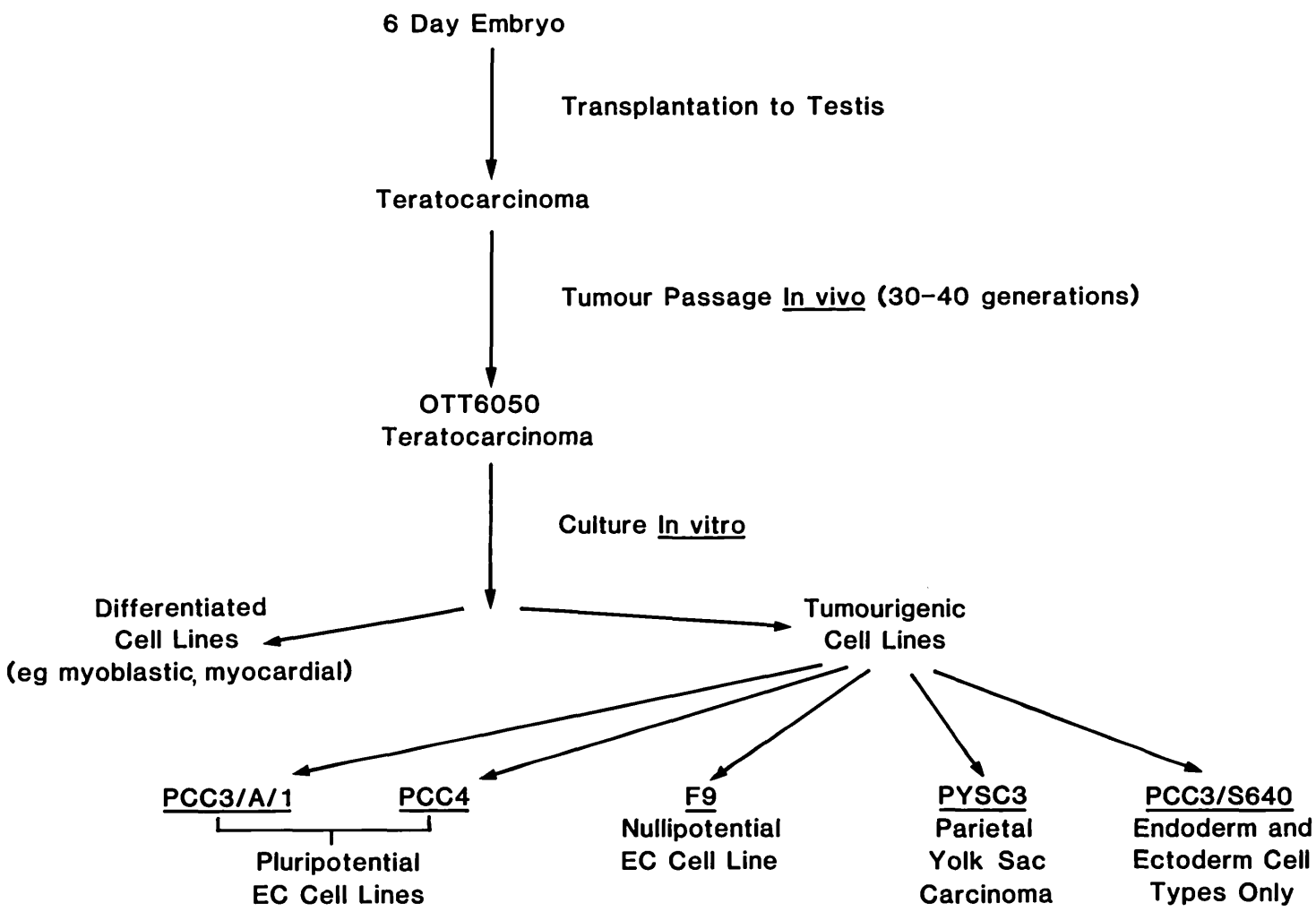
One such transplantable teratocarcinoma was OTT6050 (Stevens,

1970). As shown in figure 1.2, this teratocarcinoma was derived by transplantation of an early 6 day mouse embryo to the testis of a histocompatible host and was established as a full teratocarcinoma by serial transplantation for 30-40 generations. This teratocarcinoma contained a variety of differentiated cell types characteristic of all three germ layers in addition to cells of undifferentiated, embryonic morphology (Stevens, 1970). Subsequent *in vitro* culture of pieces of this teratocarcinoma led to the establishment of a number of clonal cell lines, each with distinct characteristics (Bernstine *et al.*, 1973; Jakob *et al.*, 1973; Boon *et al.*, 1974; fig. 1.2).

Some were non-tumourigenic, representing some of the differentiated cell types found within the teratocarcinoma (Boon *et al.*, 1974). Others were found to be tumourigenic when single cells were transplanted to histocompatible hosts. Of these, some such as PYSC3 and PCC3/S640 (fig. 1.2) were of differentiated morphology and gave rise to tumours containing only differentiated cell types. Others such as F9 (Bernstine *et al.*, 1973) and PCC3 and PCC4 (Jakob *et al.*, 1973) were of undifferentiated embryonic morphology. PCC3 and PCC4 were found to be pluripotential, in that single cells gave rise to full teratocarcinomas, containing EC cells plus differentiated cells characteristic of all three germ layers, when transplanted to ectopic sites (Jakob *et al.*, 1973). In contrast, F9 cells were found to be nullipotent in that they gave rise to tumours containing only cells of EC morphology

Figure 1.2. Origin of murine embryonal carcinoma cell lines. The origin of the murine EC cell lines used in this work is shown. All lines were derived from the teratocarcinoma OTT6050 generated by transplantation of an early 6 day mouse embryo to an ectopic site, and subsequent passage of this tumour in vivo (Stevens, 1970). In vitro culture of embryoid bodies of this teratocarcinoma resulted in the isolation of a variety of tumourigenic clonal cell lines (Bernstine et al., 1973; Jakob et al., 1973) and non-tumourigenic myoblastic and myocardial cell lines (Boon et al., 1974).

Origin of Murine Embryonal Carcinoma Cell Lines



(Bernstine et al., 1973).

More recently, pluripotential embryonic stem cell lines have been established by direct culture of blastocysts in vitro (Evans and Kaufman, 1981; Martin, 1981). Like EC cells, these embryonic stem (ES) cell lines will form teratocarcinomas when transplanted to ectopic sites. However, unlike EC cells, ES cells are derived directly from the blastocyst and have not been passaged as a tumour.

This difference is reflected in the ability of ES cells to contribute efficiently to all tissues, including the germline, when used to generate aggregation or injection chimeras with normal mouse blastocysts (Bradley et al., 1984). Whilst EC cells have been shown to contribute efficiently to all tissues in chimera formation (Brinster, 1974; Papaioannou et al., 1975; Mintz and Illemensee, 1975), they generally have abnormal karyotypes, chimeras often appear morphologically abnormal (Rossant and McBurney, 1982) and contribution to the germline is very inefficient relative to that seen with ES cells (Bradley and Robertson, 1986). The ability of ES cells to contribute efficiently to the germline in chimeric mice has important consequences with regard to the development of methods to study gene regulation and function during mouse development (section 1.4).

In addition to abnormal karyotype, the differing abilities of EC and ES cells to develop normally within the context of the embryo may well be a reflection of the different methods of isolation of such cells. EC cells have been selected in vivo,

as being unresponsive to normal differentiation signals within the teratocarcinoma. The more direct method by which ES cells are isolated circumvents this problem.

1.2.2. *In vitro* differentiation of murine embryo-derived stem cells.

In addition to differentiating *in vivo*, both EC and ES cell lines can be induced to differentiate *in vitro*, so extending the usefulness of these cell lines as *in vitro* models of the early stages of mouse embryogenesis. Many EC cell lines differentiate spontaneously, in response to culture either at too high or low density. However, a large amount of cell death and asynchrony is associated with such spontaneous differentiation and, as such, this method of differentiation is of little use.

Both EC and ES cells can be induced to differentiate *in vitro* by aggregate formation in suspension culture (Sherman, 1975; Martin and Evans, 1975; Nicolas *et al*., 1975). Within such aggregates, the outer cells alter in both morphology and biochemical properties to form extraembryonic endoderm-like cells. Further differentiation leads to the appearance of a cavity within the embryoid body to give a cystic embryoid body which can be considered analogous to the blastocyst. These embryoid bodies can subsequently be returned to monolayer culture and a wide variety of terminally-differentiated cell

types, including muscle, fibroblasts, and neural cells, appear in the culture.

The asynchrony and variety of differentiated products produced on suspension culture of EC and ES cell lines does not meet the requirement outlined earlier for pure cell populations to facilitate analyses of the regulation of gene transcription. Such problems can be alleviated, at least to some extent, by the use of chemical induction of EC cell differentiation. Bipolar compounds, such as dimethylsulphoxide (DMSO) (McBurney et al., 1982) and hexamethylenebisacetamide (HMBA) (Paulin et al., 1979) have been used to induce differentiation of EC cells, but by far the most widely used chemical inducer is all-trans retinoic acid.

The murine EC cell line F9 (Bernstine et al., 1973) is one of the most widely used EC cell lines. The low spontaneous rate of differentiation of this cell line makes it extremely useful in the analysis of the regulation of gene expression in early embryonic stem cells (section 1.3). Suspension culture of F9 cells in the presence of retinoic acid results in the formation of embryoid bodies, the surface cells of which resemble visceral endoderm by both morphological and biochemical criteria (Hogan et al., 1981). In contrast, treatment of F9 monolayers with retinoic acid plus cyclic AMP generates cells resembling parietal endoderm (Strickland et al., 1980). Thus, despite the absence of differentiation of F9 cells in vivo (section 1.2.1), manipulation of this cell line in vitro provides a reasonable model for the differentiation

Figure 1.3. *In vitro* differentiation of murine EC and ES cells.

Manipulations used to induce differentiation of EC and ES cells used in this work are shown. Treatment of monolayers of the pluripotential EC cell lines PCC3 and PCC4 (Jakob *et al.*, 1973) with retinoic acid (RA) results in the appearance of a homogeneous population of differentiated cells of fibroblast (Watson, 1986) or epithelial-like cells (Reith and Cooper, unpublished data), respectively.

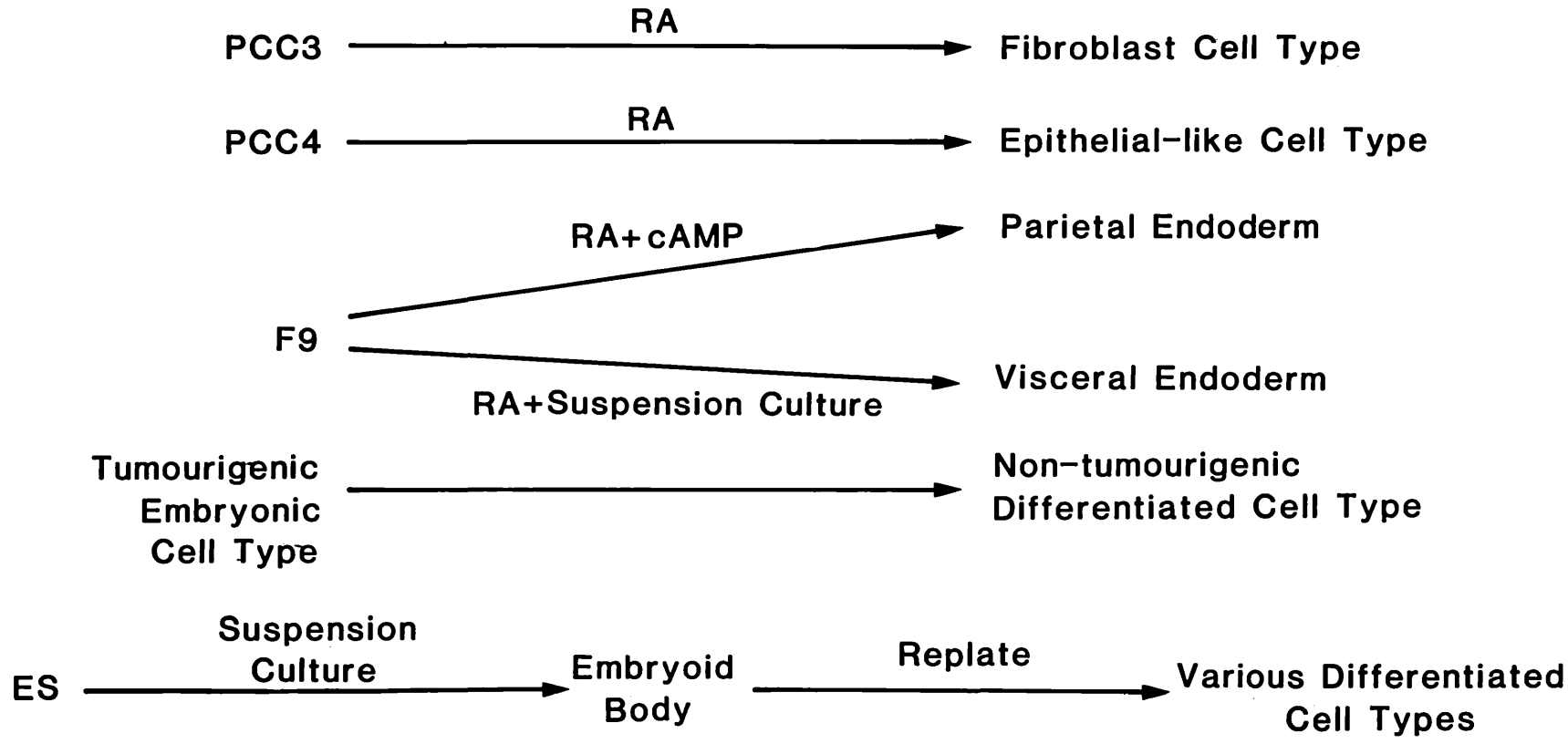
Treatment of F9 monolayers with retinoic acid plus cyclic AMP generates parietal endoderm-like cells (Strickland *et al.*, 1980), whilst suspension culture in the presence of retinoic acid generates embryoid bodies containing a high proportion of visceral endoderm cells (Hogan *et al.*, 1981).

In vitro differentiation of ES cells is achieved by suspension culture to generate embryoid bodies, prior to culture on tissue culture plastic. This generates a wide variety of terminally differentiated cell types.

In general, differentiation of embryo-derived stem cells involves a change of a rapidly proliferating, tumourigenic cell type to a non-tumourigenic, slow growing, terminally differentiated cell population.

For details of culture conditions, see section 2.3.

In vitro Differentiation of Murine Embryonal Carcinoma and Embryonic Stem Cells



of embryonic stem cells into either visceral or parietal endoderm cell types, a situation analogous to that of the formation of these cell types from the ICM *in vivo* (Hogan *et al.*, 1983; fig. 1.3).

In vitro differentiation of pluripotential EC cell lines can also be achieved by chemical induction with retinoic acid (fig. 1.3). From morphological and immunocytochemical staining studies, it was established that treatment of PCC3 monolayers with retinoic acid can give rise to a homogeneous population of differentiated fibroblast-like cells (Watson, 1986). Retinoic acid treatment of PCC4 monolayers generates a homogeneous cell population (Jetten *et al.*, 1979) which, solely on the basis of morphology, is classified as an "epithelial-like" cell type (Reith and Cooper, unpublished data).

Treatment of the murine EC cell line P19 (McBurney and Rogers, 1982) with retinoic acid leads to the appearance of a differentiated cell population containing a high proportion of neuronal cell types (Jones-Villeneuve *et al.*, 1983), whilst suspension culture of the same cell line in the presence of DMSO generates cultures containing a high proportion of skeletal and cardiac muscle cell types (McBurney *et al.*, 1982).

From the effect of retinoic acid on the various EC cell lines discussed it is apparent that, whilst this chemical stimulates EC cell differentiation, it does not influence the direction of differentiation, this being an intrinsic property of the

particular EC cell line. This, along with the finding that retinoic acid gradients are involved in the specification of anterior-posterior pattern in the developing chick limb bud (Thaller and Eichele, 1987) suggests that retinoic acid may serve as a natural morphogen in vertebrate development.

As discussed later (section 1.3), retinoic acid treatment of EC cells leads to striking alterations in gene expression patterns and the study of the molecular mechanisms by which retinoic acid affects differentiation and development is an active area of research.

There is a specific cellular retinoic acid binding protein CRABP (Jetten and Jetten, 1979; Barkai and Sherman, 1987). Some, but not all, mutant EC cell lines which do not differentiate in response to retinoic acid, are deficient in CRABP activity (Schindler et al., 1981; McCue et al., 1983; Wang and Gudas, 1984). It has been proposed that CRABP is responsible for the observed nuclear localisation of retinoic acid in F9 cells (Takase et al., 1986). Recently, the retinoic acid receptor (RAR) has been identified as a member of the steroid/thyroid hormone family of nuclear receptors (Giguere et al., 1987; Petkovich et al., 1987). The availability of this protein will greatly facilitate efforts designed to dissect the pathway by which retinoic acid brings about changes in cell phenotypes.

It is clear that embryo-derived stem cells provide a unique tool for the analysis of early differentiation events in the mouse embryo. The EC cell system has been used for immunological (Jacob, 1979; Solter and Knowles, 1979; Gooi et al., 1981; Stinnakre et al., 1981; Hyafil et al., 1983), biochemical (Dewey et al., 1978; Lovell-Badge and Evans, 1980; Morange et al., 1984) and gene regulation (section 1.3) analyses. As discussed later (section 1.4), ES cells are proving useful in the development of techniques to study gene function and regulation in vivo during mouse embryogenesis.

The murine EC cell system can be used to satisfy the three requirements given at the outset for the study of transcriptional mechanisms of gene regulation in the early mouse embryo. However, although discrete homogeneous populations of differentiated cell types can be derived from in vitro differentiation of EC cell cultures, great care should be exercised in extrapolating the experimental findings to differentiation occurring during normal mouse embryogenesis. For this reason, whenever it is possible, information obtained from the study of EC cells should be confirmed in vivo.

Section 1.3. Gene regulation in murine embryonal carcinoma cells.

1.3.1. Genes regulated in murine EC cells.

The availability of embryo-derived stem cell lines greatly facilitates the study of the transcriptional regulatory mechanisms which operate in early mouse embryogenesis. An absolute requirement for such studies are cloned genes which exhibit transcriptional regulation on EC cell differentiation. That a given gene is differentially expressed between two cell types can be established by Northern blot and transcriptional run-on analyses, as described in section 3.1. A number of genes are known to show differential expression patterns within the murine EC cell system.

Mouse EC cells have been found to be refractive to productive infection by exogenous type C retroviruses, such as murine leukemia virus, MuLV, whilst differentiated cells are permissive (Teich et al., 1977). Although virus adsorption and penetration occurred in EC cells, no viral transcription was detectable (Teich et al., 1977). A similar observation was made with the inability of DNA tumour viruses, SV40 and polyoma, to infect murine EC cells (Swartzendruber and Lehman, 1975). This restriction was subsequently shown to operate at the level of accumulation of viral mRNA (Segal and Khoury, 1979; Linnenbach et al., 1980). Similarly, the human EC cell line Tera-2 (Andrews et al., 1984) is non-permissive for human

cytomegalovirus, but becomes permissive following retinoic acid-induced differentiation of this cell line (La Femina and Hayward, 1986).

A contrasting pattern of restriction is seen with the human adenovirus type 5 (Ad5) mutant, dl312, in murine EC cells (Imperiale et al., 1984). Whilst expression of early genes of wild type Ad5 was seen in both EC cells and differentiated derivatives, early viral transcription of the E1a mutant dl312 occurred only in F9 EC cells, suggesting the presence of an E1a-like activity in EC cells but not differentiated cell types (Imperiale et al., 1984).

As discussed later (section 1.3.2), exogenous viral regulatory elements have been used extensively in the study of transcriptional gene regulation in EC cells. However, such viral elements are not normally part of the mouse genome and play no role in development. In contrast to exogenous type C retroviruses, endogenous type A retroviral particles (IAPs) are seen in early preimplantation embryos (Kelly and Condamine, 1982). Moreover, IAP RNA is specifically expressed between the late 2 cell stage and the blastocyst stage in normal mouse development (Piko et al., 1984). Expression of IAP genomes has been observed in EC cell lines PCC3 and PCC6 and is down regulated on in vitro differentiation of these EC cell lines (Hojman-Montes de Oca et al., 1983). However, the role, if any, of IAPs in mouse development is unclear.

In order to validate models of transcriptional regulation in early embryos, based on analysis of viral regulatory elements,

it is essential to study the regulation of a genuine low-copy cellular gene which is regulated during normal mouse development. Given the experimental constraints imposed on transcriptional regulation studies, such genes must also be regulated within the EC cell system. A number of genes, isolated by a variety of techniques have been shown to exhibit differential regulation in a manner analogous to that seen with viral regulatory elements.

A number of murine homeobox-containing genes have been shown to exhibit differential regulation within the EC cell system. The homeobox was originally identified as a 180 base pair nucleotide sequence encoding a highly basic peptide sequence found in the homeotic genes Antp, Ubx and ftz of Drosophila melanogaster (McGinnis et al., 1984a). The homeobox was subsequently found in a number of other genes concerned with the specification of the body plan of Drosophila melanogaster (reviewed by Gehring, 1985; Akam, 1987), including the pair rule/segment polarity gene engrailed (Fjose et al., 1985).

The homeobox sequence was found to be conserved in a variety of other species, including mouse and man (McGinnis et al., 1984b), and a number of murine homeobox-containing genes have been cloned on the basis of homology to Drosophila homeobox probes. A standardised nomenclature for murine homeobox genes has been adopted (Martin et al., 1987).

The Hox-1 cluster is found on mouse chromosome 6 and the Hox-2 cluster on mouse chromosome 11 (Martin et al., 1987). A Hox-3 locus has been mapped to mouse chromosome 15 (Awgulewitsch et

al., 1986) and a Hox-4 locus is situated on mouse chromosome 12 (Martin et al., 1987). Two murine genes, En-1 and En-2 have been cloned by means of homology with the Drosophila engrailed gene and mapped to mouse chromosomes 1 and 5 respectively (Joyner et al., 1985; Joyner and Martin, 1987).

Steady-state levels of Hox-1.1 and Hox-3.1 mRNA species have been observed to increase transiently, and with different kinetics, on retinoic acid induced differentiation of F9 cells (Colberg-Poley et al., 1985a; Brier et al., 1986). Expression of Hox-1.6 is also induced on differentiation of F9 cells to parietal endoderm (Baron et al., 1987). Induction of Hox-2.1 expression has been detected on retinoic acid induced differentiation of a number of murine EC cell lines (Deschamps et al., 1987) and the expression of En-1 increases on in vitro differentiation of the EC cell line PSA1 (Joyner et al., 1985; Joyner and Martin, 1987). Differential expression of murine homeobox genes has also been observed in normal mouse development (section 5.1.2).

A number of genes involved in growth control are also known to show differential expression in the murine EC cell system, particularly many proto-oncogenes (Adamson, 1987). Expression of int-2 has been observed to increase on in vitro differentiation of F9 cells to parietal endoderm, and in the endodermal cells of embryoid bodies derived from PSA1 EC cells (Jakobovits et al., 1986). A three-fold transient induction of c-fos expression was detected during differentiation of F9 cells to parietal endoderm (Mason et al., 1985), whilst

expression of both N-myc and c-src has been observed to decrease on F9 differentiation to parietal endoderm (Sejersen et al., 1986). In contrast, an increase in steady-state levels of both N-myc and c-src was observed on in vitro differentiation of PCC7 EC cells to a neuronal cell type (Sejersen et al., 1986). Higher levels of c-Ki-ras, c-myc and c-fos mRNA species were detected in PCC4 cells, relative to those in a terminally differentiated myocardial cell line isolated from the OTT6050 teratocarcinoma (Vilette et al., 1985; fig. 1.2). The high c-Ki-ras RNA levels may be attributable to a 10-20 fold amplification of this gene in PCC4 cells (Vilette et al., 1985).

Steady-state levels of c-myc and p53 have been observed to decrease markedly on differentiation of F9 cells to parietal endoderm (Dony et al., 1985; Sejersen et al., 1985). However, this has been shown to operate at a post-transcriptional level (Dony et al., 1985), and there is evidence to indicate that, at least in the case of c-myc, such regulation is related to the decreased proliferative activity induced by retinoic acid (Dean et al., 1986).

This latter case illustrates how great care must be exercised in interpretation of the observed patterns of down regulation of gene expression on in vitro differentiation of EC cells. The loss of the transformed phenotype and decreased proliferation associated with EC cell differentiation (fig.1.3), complicates analysis of any developmental regulation of genes within the EC cell system.

A third source of genes regulated in murine EC cells has come from the approach of differential screening of cDNA libraries. The ability to generate homogeneous cell populations on *in vitro* differentiation of EC cells (section 1.2.2) facilitates the use of this technique to isolate cDNA clones representing mRNA species, the steady-state levels of which alter on EC cell differentiation. This method not only ensures a high probability of isolating genes differentially active in a cell system amenable to experimental manipulation, but also provides an opportunity to isolate novel genes.

Using this approach it has proved possible to isolate cDNA clones representing previously characterised genes, such as laminin and type IV collagen, the expression of which increases on differentiation of F9 cells to parietal endoderm (Kurkinen *et al.*, 1983; Wang and Gudas, 1983). In addition, a number of novel genes activated (Wang *et al.*, 1985; Howe and Overton, 1986; Mason *et al.*, 1986) or repressed (Levine *et al.*, 1984; Stacey and Evans, 1984; Gorman *et al.*, 1985a; Watson, 1986) on differentiation of EC cells, or differentially active between different EC cell lines (Brulet *et al.*, 1983) have been isolated.

As discussed later (section 3.1), additional analyses of such differentially regulated clones are necessary in order to establish which, if any, are suitable for studies concerning gene regulation and function during mouse embryogenesis.

1.3.2. Transcriptional regulation of gene expression in murine EC cells.

Given the availability of a gene showing differential regulation in the EC cell system, a number of techniques are available by which the cis-acting regulatory elements of that gene can be determined. Subsequently, it is possible to characterise the trans-acting factors which interact with such nucleotide sequences, facilitating the analysis of the molecular mechanisms by which differential gene expression is achieved.

In the case of cellular genes, transgenesis offers a useful approach to the determination of the cis-elements required for the observed in vivo expression pattern of the gene, and can also be used to investigate the function of a novel gene (section 1.4). Determination of cis-regulatory elements demands the availability of the cloned locus for the gene of interest and so, in the case of genes isolated as cDNA clones, an essential first step is the isolation of genomic clones corresponding to the gene of interest. Approaches to the isolation of genomic clones are discussed in section 4.1.

In addition to the availability of genomic clones, techniques must be available which not only allow particular gene constructs to be efficiently introduced into EC cells, but also facilitate assay of promoter activity. A commonly used approach to such a problem is the use of the bacterial chloramphenicol acetyl transferase (CAT) gene as a marker

gene for promoter activity (Gorman et al., 1982). The CAT enzyme activity in crude cell extracts can be readily assayed, in a quantitative manner, by measurement of the acetylation of ^{14}C radiolabelled chloramphenicol (Gorman et al., 1982). Alternatively, it is possible to modify the coding sequence of the gene under investigation to distinguish transcripts of the exogenous gene from those derived from the endogenous locus (Scott et al., 1984).

Introduction of gene constructs into EC cells can be carried out by efficient methods of calcium phosphate-mediated DNA transfection (Gorman et al., 1984) or by infection with recombinant retroviral vectors (Rubenstein et al., 1984; Wagner et al., 1985). However, the ease of DNA transfections and various difficulties in stability and gene expression with retroviral vectors (section 1.4) have resulted in the former method being used predominantly in the study of gene regulation in EC cells. As discussed later (section 1.4), retroviral vectors are proving useful for in vivo studies of gene function and regulation.

Whilst DNA transfection experiments can be used to determine cis-regulatory sequences, further analysis of transcriptional regulatory mechanisms demands the characterisation of trans-acting factors which interact with regulatory DNA sequences. The availability of both nuclear (Dignam et al., 1983) and whole cell (Manley et al., 1980) in vitro transcription extracts, which mimic the expression pattern of the gene of interest seen in in vitro cell culture systems, has greatly

facilitated the characterisation of such factors. As discussed below, this approach, in combination with the use of gel retardation assays (Strauss and Varshavsky, 1984), has been successfully applied to the analysis of transcriptional gene regulation in EC cells.

Many of the current concepts of transcriptional gene regulation in early mouse embryogenesis have developed from the study of regulated viral promoter/enhancer elements. As mentioned earlier (section 1.3.1), murine EC cells are refractive to productive infection by many exogenous type C retroviruses and SV40 and polyoma DNA tumour viruses.

A number of polyoma mutants (Fujimura et al., 1981; Sekikawa and Levine, 1981; Vasseur et al., 1982) have been isolated which can productively infect EC cells. Transient CAT assays have shown that mutant polyoma transcriptional regulatory sequences are highly active in EC cells relative to those of wild type polyoma (Linney and Donerly, 1983). For one such polyoma mutant it has been found that this acquisition of transcriptional activity is a consequence of a single point mutation within one of the two core enhancer elements in polyoma (Herbommel et al., 1984), such that the mutant enhancer works more efficiently in both undifferentiated and differentiated cell types. In contrast, the other enhancer element, in both wild type and mutant viral genomes, showed little activity in EC cells (Herbommel et al., 1984).

Transient CAT assays have been used to investigate the inefficient expression of SV40 in murine EC cells. It has

been shown that the transcriptional activity of the SV40 early promoter occurs in an enhancer-independent manner in EC cells (Gorman et al., 1985b). However, the enhancer is required for the increased activity seen in differentiated EC cells (Sleigh and Lockett, 1985). The inactivity of murine leukemia virus (MuLV), or the closely related murine sarcoma virus (MSV), in EC cells can be alleviated by deletion of enhancer elements (Linney et al., 1984; Gorman et al., 1985b). Moreover, replacement of the SV40 enhancer with the MSV enhancer results in inactivation of the SV40 promoter (Gorman et al., 1985b), suggesting that EC cells contain trans-acting factor(s) which act via enhancer elements to reduce expression of specific genes. Recently. gel retardation assays, using MuLV LTR sequences have led to the isolation of an embryonal promoter binding factor (EPBF) found in extracts of EC cells but not differentiated cell types (Flamant et al., 1987).

A mutant of MSV, myeloproliferative sarcoma virus (MPSV), unlike the parental virus, can efficiently direct gene expression in hemopoietic and F9 EC cells (Seliger et al., 1985; Franz et al., 1986). Several point mutations within the U3 region of the MPSV LTR have been shown to contribute to this expanded host-range (Stocking et al., 1985, 1986). However, gene expression directed by the MPSV LTR is 2-3 orders of magnitude lower in the pluripotential EC cell line PCC4 relative to F9 cells (Franz et al., 1986). Selection for neo^r PCC4 colonies following infection with an MPSV-neo recombinant retrovirus, led to the isolation of another mutant

virus, PCMV, the LTR of which directs equivalent high levels of gene expression in both F9 and PCC4 cells (Franz et al., 1986; Hilberg et al., 1987). Analysis of the PCMV genome indicates that the expanded host-range of this virus is a function of further nucleotide changes within the MPSV LTR (Hilberg et al., 1987).

There is also evidence to suggest that additional mechanisms may operate on the MuLV LTR to account for the inefficient transcriptional activity of this element in EC cells.

Selection for recombinant MuLV genomes transcriptionally active in F9 cells led to the isolation of MuLV genomes containing single point mutations in the tRNA primer binding site of the MuLV LTR (Barklis et al., 1986). Moreover, deletion analysis using DNA transfection and transient assays of gene activity indicated that the inhibition of MuLV transcription in EC cells is mediated by a nucleotide sequence which includes this tRNA primer binding site (Loh et al., 1987). However, it is unclear whether this element acts at the level of transcription or RNA stability.

In addition to the role of trans-acting factors in the regulation of viral gene expression in murine EC cells, there is evidence to suggest that methylation of specific nucleotide sequences may also be involved in the regulation of both viral and endogenous gene expression in EC cells (Stewart et al., 1982; Young and Tilghman, 1984; Jaenisch and Jahner, 1984). Whilst exogenous type C retroviruses are expressed inefficiently in EC cells, it is worthy to note that some

endogenous type A retroviral LTR elements are highly active in transient CAT assays in EC cells (Morgan et al., 1988), indicating that these viral transcriptional regulatory sequences are not subject to negative regulation in embryonic stem cells.

The ability of murine EC cells, but not differentiated cell types, to complement E1a deficiency in a human adenovirus type 5 mutant (Imperiale et al., 1984) has led to much interest in the cellular E1a-like activity in EC cells. It is known that E1a will repress SV40 and polyoma transcription, via interaction of trans-acting factors with the viral enhancer elements (Borelli et al., 1984; Velich and Ziff, 1985). Moreover, a polyoma mutant which can productively infect murine EC cells escapes such E1a-mediated repression (Hen et al., 1986). By use of in vitro transcription extracts and gel retardation assays, it has been shown that murine EC cells, but not differentiated cell types, contain a DNA binding factor which interacts specifically with an E1a-dependent adenovirus promoter (La Thangue and Rigby, 1987).

In contrast to the analysis of transcriptional regulation of viral genomes in EC cells, there is relatively little data concerning regulatory mechanisms known to operate on genuine cellular genes.

Introduction of exogenous cellular genes, such as H2-K^b (Rosenthal et al., 1984), δ-crystallin (Kondoh et al., 1984) or human cardiac actin (Rudnicki et al., 1988) has shown that, under the regulation of their own promoters, these genes are

not expressed in EC cells. However, expression of the exogenous genes was seen on induction of differentiation in a manner similar to that seen for the endogenous gene. This allows the EC cell system to be utilised to dissect the promoter regions of these genes in order to determine the cis-regulatory sequences responsible for the observed expression pattern.

The best example of such an approach is provided by the analysis of α-fetoprotein (AFP) expression in the F9 EC cell system (Scott et al., 1984). AFP expression is known to be induced on in vitro differentiation of F9 EC cells to visceral endoderm (Hogan et al., 1981). Various deletion constructs of the AFP gene were introduced into F9 cells by transfection. By using a deleted coding sequence, transcripts from the exogenous AFP gene construct could be distinguished from those of the endogenous gene.

From such analyses, it was possible to determine the cis-elements required to mimic the qualitative, but not quantitative, induction of AFP expression on visceral endoderm formation. Like the endogenous AFP gene, no expression of the exogenous genes was detected in F9 EC cells, or F9 cells induced to differentiate to parietal endoderm (Scott et al., 1984).

In conjunction with further promoter deletion analyses in the human hepatoma cell line hepG2 (Godbout et al., 1986) and transgenic experiments (Krumlauf et al., 1985; Hammer et al., 1987; section 1.4), three enhancer elements and one proximal

promoter element lying within 7kb from the transcriptional start of the AFP gene, were shown to mediate all developmental and tissue-specific patterns of AFP expression in the mouse (section 1.4).

Recently, further transfection assays in the F9 EC cell system and hepG2 cells have indicated that there is an absolute requirement for AFP enhancer elements in the activation of AFP expression on visceral endoderm formation (Vogt et al., 1988). This enhancer requirement is in addition to that needed for high level AFP expression in differentiated cells. As discussed later (section 1.4), the study of the regulation of AFP expression illustrates the importance of germline manipulation in the study of the regulation of expression of cellular genes during mouse embryogenesis.

It should also be noted that much of the analysis of gene regulation in EC cells has focused on the mechanisms which operate on genes expressed in differentiated, but not undifferentiated, EC cells. A cellular E1a-like activity, implied from the growth of the E1a-deficient mutant dl312 in F9 cells (Imperiale et al., 1984), can facilitate expression of viral and, presumably, cellular genes in EC cells. The isolation of cellular genes expressed in EC cells but not in differentiated cell types would allow such models of transcriptional regulation of gene expression to be tested.

Section 1.4. Analysis of gene function and regulation during mouse embryogenesis.

Given the availability of a novel gene regulated in EC cells, it is possible to investigate the regulation of that gene within the EC cell system using the methods outlined in section 1.3. However, as discussed earlier (section 1.2) any information gained from the study of gene function and regulation in embryo-derived stem cells must be confirmed by analysis of that gene in normal mouse development. Particularly, regulated gene expression must be seen during mouse development, as well as in the EC cell system.

Experimental approaches to the determination of gene expression patterns during mouse development are discussed in chapters 3 and 5. Such analyses may reveal additional cell type-specific expression patterns for which appropriate *in vitro* model cell culture systems are available. Should a novel gene be under investigation, the nature of the gene product can be determined by amino acid sequence and biochemical analyses. Approaches to such analyses are discussed in chapter 6. In combination with the expression studies, such information will allow predictions to be made concerning potential functions of that gene during development. Any further analyses of both gene function and regulation during mouse embryogenesis are based upon the ability to manipulate the mouse genome.

The ability to introduce new genetic information into the

mouse germline, in an experimentally manipulatable manner, is based upon the development of *in vitro* embryo culture methods. Early embryos can be isolated and maintained for short periods *in vitro* before being returned to the uterus, with no apparent interference with normal development (Brinster, 1972).

Microinjection of DNA into the fertilised pronucleus is by far the most prevalent technique currently used to introduce foreign genes into the mouse germline. Following the first success of this technique (Gordon *et al.*, 1980) it has developed rapidly, and many transgenic lines of mice have been generated (reviewed by Palmiter and Brinster, 1986). The injected DNA integrates into the genome apparently at random and often in tandem arrays. Once integrated, it is generally inherited in a typical Mendelian manner. Although technically demanding, this approach can be used to introduce DNA from any source, and of any size, into the mouse germline. Whilst, in most, but not all, cases the transgene can show appropriate qualitative patterns of expression, transgene activity is often well below that seen for the endogenous gene and is influenced by the site of integration (reviewed by Palmiter and Brinster, 1986). Recent advances in the use of appropriate *cis*-regulatory sequences (Grosveld *et al.*, 1987) may overcome this problem.

Another route to the germline is offered by retroviral infection of preimplantation mouse embryos (Jaenisch, 1976). The genome of Moloney murine leukemia virus (MuLV) has been

manipulated to facilitate its use as a vector for the introduction of novel DNA sequences into cells (Mann et al., 1983). The advantages of this approach are the ease of embryo manipulations involved, the high efficiency of infection and the generation of single copy integrations in an apparently random manner (Wagner and Stewart, 1986). However, there is a size limit of 6-8kb in the amount of DNA that can be included in the vector and recombinant retroviral vectors can be unstable and often undergo genetic rearrangements (Joyner and Bernstein, 1983). Moreover, integration occurs after cell division has commenced in the embryo, generating mosaic founder animals, and so additional breeding is required in order to establish transgenic lines.

Whilst introduction of novel sequences into the germline has been achieved by this method, the block to MuLV LTR activity in early embryonic cells (section 1.3) prevented efficient expression of integrated genes either in embryonic or adult tissues (Huszar et al., 1985; Jahner et al., 1985; Van der Putten et al., 1985; Rubenstein et al., 1986). The use of an internal promoter has allowed transgene expression in mice via embryo infection with retroviral vectors (Stewart et al., 1987). However, the inclusion of internal promoters limits, still further, the capacity of retroviral vectors.

An additional approach to the introduction of stably inherited alterations into the mouse germline is the use of genetically manipulated embryo-derived stem cell lines in chimera formation with normal mouse blastocysts. Germline

chimerism can be achieved with high efficiency using ES cells (Bradley et al., 1984). This method is particularly appropriate in the present studies, given the usefulness of embryo-derived stem cell lines in the analysis of gene regulation and function in the early mouse embryo (section 1.2; section 1.3).

The attraction to the use of ES cells as a route to the germline lies in the ability to screen or select the stem cell clone containing the desired alteration prior to introduction to the germline. For example, selection for neomycin resistance, following infection with an LTR-driven neo^r retroviral vector, led to the isolation of EC and ES cell clones expressing the bacterial neo gene (Stewart et al., 1985). Chimeric mice generated from such clones were subsequently found to express neo in all tissues (Stewart et al., 1985). The introduction of the neomycin gene into the mouse germline, via selection of ES cells in vitro (Gossler et al., 1986), has demonstrated the feasibility of this approach to germline manipulation. However, this strategy is limited by the design of the screening/selection method to be used, as well as the difficulty in maintaining the normal diploid chromosome complement of ES cells.

Transgenesis can be used to investigate both gene regulation and function in the early mouse embryo. Determination of cis-acting transcriptional regulatory elements by this method is exemplified by the analysis of α-fetoprotein (AFP) expression in transgenic mice. This example also illustrates how both in

vitro cell lines and transgenesis can be combined to characterise cis-acting transcriptional regulatory elements.

As discussed earlier (section 1.3), introduction of various deletion constructs of the AFP gene into the F9 EC cell system facilitated the characterisation of the cis-elements responsible for transcriptional regulation of this gene in visceral endoderm cells (Scott et al., 1984). However, expression of AFP was also seen in fetal liver and gut, but AFP RNA was not detectable in neonatal liver and gut. The absence of steady-state levels of AFP RNA in neonatal liver is known to operate at the level of transcription (Tilghman and Belayew, 1982). Induction of AFP was also seen in regenerating mouse liver (Belayew and Tilghman, 1982). In the absence of any suitable in vitro cell system for such tissues, transgenesis offered the means to study the mechanism of regulation of AFP.

The generation of transgenic mice bearing AFP minigenes demonstrated that the tissue-specific and developmental expression of AFP could be mediated, qualitatively, by the use of constructs containing 7kb of DNA sequence upstream of the transcriptional start of this gene (Krumlauf et al., 1985).

Transient assays of AFP deletion constructs in the human hepatoma cell line hepG2 characterised 3 enhancer elements, and one additional cell type-specific regulatory element, lying within the 7kb regulatory element of the AFP gene (Godbout et al., 1986). Further transgenic experiments indicated that at least one of the enhancer elements is

necessary to establish AFP expression in visceral endoderm, fetal liver and fetal gut and allow AFP induction in regenerating liver (Hammer et al., 1987). In contrast to the studies with in vitro cell lines, this work suggested that the 3 enhancer elements were not equivalent and may be responsible for the differing levels of AFP seen in different tissues.

Transgenesis using fusion gene constructs has also been used to investigate the function of a given gene product in development. A case in point is the analysis of the phenotypic consequences of deregulated expression of c-fos in transgenic mice (Ruther et al., 1987).

In these mice, the c-fos structural gene, deleted in 3' regulatory sequences, was expressed from the metallothionein promoter. c-fos expression was detected in a wide variety of tissues in the transgenic mice, but only bone development was found to be abnormal. Specifically, swellings were detected at the epiphysal surfaces of developing bones (Ruther et al., 1987). This result is compatible with the known expression pattern of c-fos during mouse development (Dony and Gruss, 1987b), as well as the origin of v-fos in oncogenic retroviruses known to cause osteosarcomas (Muller, 1986). However, unlike v-fos, deregulated c-fos expression in transgenic mice did not result in tumour formation (Ruther et al., 1987).

Analysis of in vivo functions of a gene can also be investigated by analysis of the phenotypic consequences of experimentally induced loss of gene function in transgenic

animals.

Replication competent MuLV, and derivatives, have been used to infect both preimplantation and post-implantation mouse embryos in attempts to utilise retroviruses as insertional mutagens to study gene function and regulation during mouse development (Jaenisch et al., 1983; Soriano et al., 1987). Of 44 transgenic strains bearing a total of 48 proviral integrations, two, Mov 13 (Jaenisch et al., 1983) and Mov 34 (Soriano et al., 1987), were found to carry recessive lethal mutations.

The Mov 13 mutation was lethal at 12-14 days gestation (Jaenisch et al., 1983) and proviral integration was found to be within the first intron of the α1(I) collagen gene, such that transcription of this gene was inhibited (Schnieke et al., 1983; Harbers et al., 1984; Hartung et al., 1986). This mutant strain has facilitated studies of the role of α1(I) collagen in development (Lohler et al., 1984; Kratochwil et al., 1986; Schnieke et al., 1987). The Mov 34 strain carries a recessive mutation lethal just after implantation of the blastocyst (Soriano et al., 1987). Proviral integration is within the 5′ end of a ubiquitously transcribed gene of novel nucleotide sequence (Soriano et al., 1987). It should also be noted that the generation of transgenic mice by microinjection of DNA has also produced a number of developmental mutations (Wagner et al., 1983; Palmiter et al., 1984; Mark et al., 1985; Woychik et al., 1985; Overbeek et al., 1986; reviewed by Palmiter and Brinster, 1986).

In the above cases, the loss of gene function was achieved by random integration of exogenous DNA. A more precise method for analysis of the consequences of the loss of gene function requires the ability to specifically disrupt the function of a chosen gene product of interest. Such a directed loss of gene function can, in theory, be brought about by introduction of anti-sense RNA (Izant and Weintraub, 1984), over-production of a mutant polypeptide (Herskowitz, 1987) or by means of homologous recombination, so as to disrupt the locus of interest (Smithies et al., 1985).

Anti-sense gene regulation in transgenic mice offers a potentially powerful method for the analysis of gene function during mouse embryogenesis. In Drosophila melanogaster, anti-sense gene constructs have been used to generate phenocopies of Kruppel mutants (Rosenberg et al., 1985), whilst in Dictyostelium, anti-sense RNA has been used to inactivate myosin heavy chain gene expression (Knecht and Loomis, 1987). In higher eukaryotes, anti-sense c-fos mRNA inhibits PDGF-induced stimulation of mouse 3T3 cell proliferation (Nishikura and Murray, 1987) and anti-sense c-myc oligonucleotides inhibit the entry of human T lymphocytes into S-phase of the cell cycle (Heikkila et al., 1987). Transgenic mice have also been generated which express anti-sense HPRT mRNA (Munir et al., 1987; Stout and Caskey, 1987). In such mice, a 50% reduction in HPRT protein was observed, but no phenotypic effects were apparent (T. Caskey, pers. comm.). The ratio of anti-sense:sense mRNA species required to give this observed

reduction in protein level is in the order of 40:1, and so powerful transcriptional regulatory elements are required to produce appropriate amounts of anti-sense RNA.

Transgenesis using gene constructs which over-express a specifically mutated polypeptide also offers a potentially powerful method for analysis of gene function in the developing embryo (Herskowitz, 1987). One example of this approach is provided by the generation of transgenic mice bearing a mutated dihydrofolate reductase (DHFR) cDNA under the transcriptional regulation of an SV40 promoter (Gordon, 1986). This DHFR cDNA sequence carried a single point mutation known to raise the K_m of the DHFR enzyme and reduce dihydrofolate turnover by a factor of 20 (Haber et al., 1981). This transgene, when present as a single copy in the genome, was found to induce a dominant mutant phenotype which included stunted growth and skeletal defects (Gordon, 1986).

The ability to propagate ES cells for long periods in vitro facilitates the selection and characterisation of specific mutations which can subsequently be introduced into the germline. In vitro selection for HPRT-deficient ES cell lines has led to the generation of HPRT deficient mice via chimeras (Hooper et al., 1987; Kuehn et al., 1987). The ease of selection for, or against, HPRT activity in in vitro cell lines has also been utilised in the development of techniques for homologous recombination in ES cells (Doetschman et al., 1987; Thomas and Capecchi, 1987).

The conceptual basis for such work is provided by studies of

homologous recombination in yeast (Orr-Weaver et al., 1981). Using HPRT as a model gene, a frequency of homologous recombination of 10^{-3} of transfected cells has been achieved (Thomas and Capecchi, 1987). Given this low frequency, and the lack of selection for inactivation of most loci, sensitive screening methods will have to be employed to isolate ES clones bearing the desired mutations. In this respect, the polymerase chain reaction (PCR) method (Mullis and Faloona, 1987), in conjunction with the use of selectable marker genes such as that for neomycin resistance, may prove useful.
Once sensitive methods have been developed to detect ES clones bearing mutations in loci not so readily assayed as HPRT, it should be possible to disrupt any locus for which genomic clones are available. Such ES cells could then be used to introduce specific mutations into the germline, allowing the phenotypic consequences of a loss of gene function to be investigated at all stages of development.

From the discussion so far, it is apparent that the analysis of transcriptional regulation of gene expression in early mouse embryogenesis is dependent upon the use of embryo-derived stem cell lines. Many genes isolated from EC cells are activated at late stages in development and are expressed in terminally differentiated cell types. In order to test models of transcriptional regulation in pluripotential embryonic stem cells it is necessary to utilise an endogenous cellular gene

expressed in pluripotential cells but not in their differentiated derivatives. Genes highly active in stem cells, but strongly down-regulated on differentiation, would not only make transcriptional regulation experiments easier to design, but would prove a useful addition to the repertoire of cell type-specific promoter/enhancer elements currently available for expression of fusion genes in transgenic mice.

Analysis of the mechanisms that control mammalian development is hampered by the lack of molecularly defined embryological mutants and developmentally regulated genes. The isolation of novel genes for use in transcriptional regulation studies would allow the function of that gene in normal mouse development to be investigated.

Given these aims, a cDNA library prepared from the pluripotential EC cell line PCC3 (Jakob et al., 1973) was screened in order to isolate cDNA clones representing genes, the steady-state levels of which decrease on in vitro differentiation of this cell line (Gorman et al., 1985a; Watson, 1986; Cooper and Watson, unpublished data).

An analysis of a number of these cDNA clones was initiated to determine which, if any, would prove useful in the study of gene regulation and function during mouse development. From such an analysis (Chapter 3), two cDNA clones, GR1 and 1.6U, were considered worthy of further study.

GR1 exhibits clear down regulation on retinoic acid-induced differentiation of PCC3 cells and is also regulated during mouse development. Studies of the mechanism of transcriptional

regulation of this novel gene were initiated by the isolation of genomic clones corresponding to the GR1 locus (Raleigh et al., 1988; Chapter 4).

A detailed analysis of the expression pattern of this novel gene during mouse development was carried out by means of RNA in situ hybridisation to mouse embryos, and attempts were made to raise polyclonal antisera against the putative GR1-encoded polypeptide (Chapter 5).

The GR1 mRNA species is 6.8kb in size. The sequence of 2674 nucleotides in this transcription unit was determined. Analysis indicated that the determined sequence corresponds to 247 amino acids at the carboxy terminus of the putative GR1 polypeptide, there being at least 1.9kb of 3′ untranslated sequences in the GR1 mRNA. The carboxy terminal domain of the putative GR1 polypeptide is proline-rich and contains extensive regions of -sheet secondary structures (Chapter 6). The GR1 locus was mapped to mouse chromosome 6 on the basis of linkage to the met proto-oncogene (Chapter 6).

The 1.6U cDNA clone was found to represent the transcript(s) of a novel gene present in a number of embryonic and terminally differentiated cell types (chapter 3). These transcripts have been shown to be the products of a highly active promoter (A.R. Cooper, unpublished data). As such, the promoter elements of 1.6U may prove to be useful in obtaining high constitutive levels of gene expression in a wide variety of tissues in transgenic mice. To this end, genomic clones of the 1.6U gene(s) were isolated (Raleigh et al., 1988; Chapter

4) in order to characterise the constitutively active transcriptional regulatory elements of this gene.

Chapter 2. Materials and Methods.

Section 2.1. Bacterial strains.

Strains of E.coli used in this work are given in Table 2.1. Single colonies of E.coli strains picked from L-agar plates were used to inoculate 10ml cultures of L-broth or psi-broth (Table 2.2). Colonies of E.coli carrying plasmids or cosmids were picked from L-agar plates containing the appropriate antibiotic and inoculated into 10ml cultures of L-broth or psi-broth supplemented with that antibiotic. E.coli strain JM103 was grown on minimal agar (Table 2.2).

Following incubation at 37°C with vigorous aeration for 12-18 hours, aliquots were made 40% (v/v) in glycerol and stored at -70°C.

For the preparation of galactosidase-fusion proteins a single ampicillin resistant colony bearing the appropriate plasmid was inoculated into 2-4ml L-broth containing 100ug/ml ampicillin (Sigma), 500uM isopropyl β-galactoside (Pharmacia) and incubated at 27°C or 37°C for 16 hours. Isolation of proteins was as described in section 2.12.

Section 2.2. Bacteriophage strains.

Bacteriophage strains used in this study are given in Table 2.3. Single plaques were picked with a sterile glass pasteur pipette into 0.5ml SM buffer (Table 2.2) and stored at 4°C. Plating cells were prepared by inoculating 50ml L-broth,

Table 2.1.

Bacterial strains used for growth of plasmids and bacteriophage.

Strain	Genotype	Source	Reference
E.coli C600	F^-,*thi*-1,*thr*-1,*leu*b6, *lac*Y1,*ton*A21,*sup*E44	C.J. Watson[1].	Appleyard 1954.
E.coli HB101	F^-,*hsd*20(r^-_B,m^-_B), *rec*A13,*ara*-14,*pro*a2, *lac*Y1,*gal*K2, *rps*L20(Sm^r),*xyl*-5, *mtl*-1,*sup*E44.	A.R. Cooper[1].	Boyer and Roulland-Dussoix, 1969.
E.coli BMH71-18	Δ(*lac*,*pro*), F'*lac*$I^qZ\Delta$M15*pro*$^+$	N. Williamson[1].	Messing *et al*., 1977.
E.coli JM103	Δ(*lac*,*pro*),*thi*, *str*A,*sup*E,*end*A *sbc*B, *hsd*R^-,F'*tra*D36,*pro*AB *lac*$I^qZ\Delta$M15.	E.J. Murray[1].	Messing *et al*., 1981.
E.coli LE392	F^-,*hsd*R514(r^-_k,m^+_k), *sup*E44,*sup*F58,*lac*Y1, *gal*K2,*gal*T22,*met*B1, *trp*R55.	Stratagene.	Murray *et al*., 1977.
E.coli P2392	P2 lysogen of LE392	Stratagene.	Stratagene pers.comm.

Table 2.1. cont.

E.coli NM621	C600 *sup*E,*hsd*R,*mcr*A, *mcr*B,*rec*D1009.	D.R.F. Leach[2]	Raleigh *et al.*, 1988.
E.coli CES200	F^-,*rec*B21,*rec*C22, *sbc*B15,*thr*1,*leu*6, *pro*A2,*his*4,*thi*1, *arg*E3,*lac*Y1,*gal*K2, *ara*14,*xyl*5,*mtl*1, *str*31,*tsx*33,thr^+, leu^+, (*svl*-*rec*A)Δ306::Tn*10*.	D.R.F. Leach[2]	Leach, pers.comm.
E.coli ED8767	*rec*A56,*sup*E,*sup*F, *hsd*S^-,R^+M^+ *met*$^-$	E.J. Murray[1]	Murray *et al.*, 1977.
E.coli ED8910	*rec*B21,*rec*C22, *sup*F,*hsd*$^-$,R^-M^+	W.J. Brammar[3]	Brammar pers.comm.

1= Department of Biochemistry, Imperial College of Science and Technology, University of London.

2= Department of Molecular Biology, University of Edinburgh.

3= Department of Biochemistry, University of Leicester.

supplemented with 0.2% (w/v) maltose (BDH), with a single colony of the appropriate strain of E.coli and incubating overnight with vigorous aeration at 37°C.

The bacteriophage stock was titred by adding 3ul of bacteriophage diluted in SM buffer (Table 2.2) to a mixture of 100ul plating cells, 100ul 10mM magnesium chloride, 10mM calcium chloride, and 100ul SM buffer. Following incubation at 37°C for 15 minutes, 3ml top agarose (Table 2.2) at 45°C were added, the mixture plated and incubated overnight at 37°C.

High titre bacteriophage stocks were made according to the method of Maniatis et al. (1982). Sufficient bacteriophage to give complete plate lysis were plated as above. Following overnight incubation plates were overlaid with 3ml SM buffer and incubated at 4°C with gentle shaking for 4-8 hours. The buffer was then removed, titred, and stored at 4°C.

M13 bacteriophage were stored as single-stranded DNA preparations as described in section 2.22.1.

Section 2.3. Mammalian cell culture.

Mammalian cell lines (Table 2.4) were grown at 37°C in an incubator gassed with 8% (v/v) carbon dioxide in air of relative humidity greater than 95%. All cell lines were grown in Dulbecco's modification of Eagle's medium (Flow laboratories) supplemented with 10% (v/v) fetal calf serum (Gibco), 500u/ml penicillin (Sigma), and 100ug/ml streptomycin

Table 2.2.

Media used for the growth of E.coli.

L-broth	: 1% (w/v) Difco tryptone, 0.5% (w/v) Difco yeast extract, 0.5% (w/v) sodium chloride, adjusted to pH7.5 with sodium hydroxide.
psi-broth	: 2% (w/v) Difco tryptone, 0.5% (w/v) Difco yeast extract, 0.05% (w/v) sodium chloride, 0.25% (w/v) magnesium sulphate, adjusted to pH7.5 with sodium hydroxide.
NZCYM-broth	: 1% (w/v) Oxoid casein hydrolysate, 0.5% (w/v) sodium chloride, 0.1% (w/v) Difco casamino acids, 0.1% (w/v) Difco yeast extract, 0.2% (w/v) magnesium sulphate, adjusted to pH7.5 with sodium hydroxide.
TY-broth	: 1.6% (w/v) Difco tryptone, 1% (w/v) Difco yeast extract, 0.5% (w/v) sodium chloride, natural pH
L-agar	: L-broth supplemented with 1.5% (w/v) Difco agar.
NZCYM-agar	: NZCYM-broth supplemented with 1.5% (w/v) Difco agar.
Top agarose	: L-broth containing 0.75% (w/v) Sigma agarose.

Table 2.2. cont.

Minimal agar : 0.6% (w/v) disodium hydrogen phosphate, 0.3% (w/v) potassium dihydrogen phosphate, 0.05% (w/v) sodium chloride, 0.1% (w/v) ammonium chloride, 0.2% (w/v) glucose, 0.025% (w/v) magnesium sulphate, 0.001% (w/v) calcium chloride, 1mM thiamine HCl, 1.5% (w/v) Difco agar, adjusted to pH7.4.

SM buffer : 100mM sodium chloride, 10mM magnesium sulphate, 50mM Tris-HCl pH7.5, 0.01% (w/v) gelatin.

Table 2.3

Bacteriophage and Plasmids.

Bacteriophage or Plasmid	Host *E.coli*	Source	Reference
M13mp8	JM103	K. O'Hare[1]	Messing and Vieira, 1982.
M13mp9	JM103	K. O'Hare[1]	
M13mp19	JM103	D. Hayward[1]	Norrander *et al.*, 1983.
λ22	C600	C.J. Watson[1]	Gorman *et al.*, 1985a.
λGR15	C600	C.J. Watson[1]	Watson, 1986.
λGR1	C600	C.J. Watson[1]	Watson, 1986.
λGR4	C600	C.J. Watson[1]	Watson, 1986.
pSP65	HB101	A.R. Cooper[1]	Melton *et al.*, 1984.
pBSM13$^+$	JM103	A. Mitcheson[1]	Stratagene pers. comm.
pBR325	HB101	A.R. Cooper[1]	Bolivar, 1978.
pUR290	BMH71-18	W. Whitfield[1]	Ruther and Muller-Hill, 1983.
pUR291	BMH71-18	W. Whitfield[1]	
pUR292	BMH71-18	W. Whitfield[1]	
pHPT4	HB101	D. Melton[2]	Konecki *et al.*, 1982.
pPE 9	HB101	A.R. Cooper[1]	Barlow *et al.*, 1984.

Table 2.3. cont.

p64B2	HB101	A.R. Cooper[1]	**
pBSBT	HB101	This work.	***

** p64B2 consists of the B2 containing 64E restriction fragment of the cDNA clone pAG64 (Brickell et al., 1983) subcloned as a EcoRI-HindIII restriction fragment into pSP64 (A.R. Cooper, pers. comm.).

*** pBSBT contains a human αtubulin cDNA clone (Cowan et al., 1983) subcloned as a 1.4kb PstI-PstI restriction fragment into pBSM13$^+$ (data not shown).

1= Department of Biochemistry, Imperial College of Science and Technology, University of London.

2= Department of Molecular Biology, University of Edinburgh.

(Sigma).

Differentiation of embryonal carcinoma cell lines was induced by the addition of $5x10^{-7}$ M retinoic acid (Sigma) for PCC3 cells and $5x10^{-5}$ M retinoic acid for PCC4 cells. Differentiation of the embryonal carcinoma cell line F9 to parietal endoderm was induced by addition of $5x10^{-8}$ M retinoic acid, $1x10^{-3}$ M cyclic AMP and $1x10^{-4}$ M IBMX (Strickland and Mahdavi, 1978; Strickland et al., 1980). Medium was replaced every two days.

Section 2.4. Animals.

Adult and newborn Balb/c mice, pregnant CBA mice (normal cycle), adult DBA2/J, and adult C57bl/6 mice, as well as half-lop rabbits were obtained from the Biological Services Division, NIMR, London.

Section 2.5. Large scale preparation of bacteriophage lambda DNA.

Bulk preparations of bacteriophage DNA were made according to the method of Dr.P.Brickell (Middx. Hospital). The chosen strain of E.coli was grown at 37°C in L-broth (Table 2.2), supplemented with 0.2% (w/v) maltose, to late log phase. To 1ml of this culture was added an equal volume of 10mM

Table 2.4

Mammalian cell culture.

Cell type	Name	Source	Reference
Murine embryonal carcinoma cell line	PCC3	C.J. Watson[1]	Jakob et al., 1973.
Murine embryonal carcinoma cell line	PCC4	C.J. Watson[1]	Jakob et al., 1973.
Murine embryonal carcinoma cell line	F9	A.R. Cooper[1]	Bernstine et al., 1973.
Murine Balb/c 3T3 fibroblasts	A31	B.I. Skene[1]	Aaronson and Todaro, 1968
Murine Balb/c 3T3 clone clone A31 fibroblasts transformed by wt830 SV40	SV3T3 C138	B.I. Skene[1]	Rigby et al., 1980.

1= Department of Biochemistry, Imperial College of Science and Technology, University of London.

magnesium chloride, 10mM calcium chloride and 1ml of an appropriate dilution of bacteriophage stock to give a multiplicity of infection of 0.1. Following incubation at 37°C for 20 minutes the mixture was added to 500ml L-broth containing 10mM magnesium chloride, prewarmed to 37°C, and incubated with vigorous aeration at 37°C for 12-16 hours.

After the addition of 2.5ml chloroform and 30g sodium chloride, bacterial debris was removed by centrifugation at 8000rpm for 10 minutes at 4°C in a Sorvall GS-3 rotor. The supernatant was transferred to a fresh container and polyethylene glycol 8000 (BDH) dissolved at room temperature to a final concentration of 100g per litre. Following incubation for 16 hours at 4°C, precipitated bacteriophage particles were collected by centrifugation at 7500rpm for 30 minutes at 4°C in a Sorvall GS-3 rotor.

The pellet was dissolved in 13ml caesium chloride solution (1.47g/ml) and transferred to an SW41 polyallomer tube (Beckman). After overnight incubation at 4°C the mixture was centrifuged at 32000rpm for 24-48 hours at 4°C in a Beckman SW41 rotor. Bacteriophage particles were collected by side puncture and caesium chloride removed by dialysis at 4°C for 2 hours against 10mM sodium chloride, 50mM Tris-HCl, pH8.0, 10mM magnesium chloride.

The bacteriophage suspension was adjusted to 20mM EDTA, pH8.0, 0.5% (w/v) sodium dodecyl sulphate, and proteinase K (Sigma) added to a final concentration of 50ug/ml. Following incubation at 65°C for 60 minutes, the solution was extracted

twice with phenol:chloroform:isoamylalcohol (50:50:1), twice with water-saturated ether, and DNA precipitated by the addition of two volumes of absolute ethanol.

DNA was collected by centrifugation at 8000rpm for 15 minutes at $4^{o}C$, the pellet washed thoroughly with 70% (v/v) ethanol, dried under vacuum, redissolved in a minimum volume of TE (10mM Tris-HCl, pH7.5, 1mM EDTA, pH8.0) and stored at $4^{o}C$.

Section 2.6. Large scale preparation of cosmid and plasmid DNA.

500ml cultures of E.coli strains bearing plasmids or cosmids were grown at $37^{o}C$ to late log phase in L-broth or psi-broth (Table 2.2) containing 100ug/ml ampicillin (Sigma). Cells were harvested by centrifugation at 7000rpm for 5 minutes at $4^{o}C$ in a Sorvall GS-3 rotor.

The resulting pellet was resuspended in 18ml 50mM glucose (Sigma), 25mM Tris-HCl, pH8.0, 10mM EDTA, pH8.0, 5mg/ml lysozyme (Sigma), and incubated at room temperature for 10 minutes. 40ml of a freshly prepared solution of 0.2M sodium hydroxide, 1% (w/v) sodium dodecyl sulphate were added, gently mixed, and the preparation incubated on ice for 5 minutes.

After the addition of 20ml precooled 3M potassium acetate, pH4.8, and subsequent incubation on ice for 15 minutes, the resulting precipitate was removed by centrifugation at 8000rpm

for 5 minutes at 4°C in a Sorvall GS-3 rotor. The supernatant was filtered through muslin and episomal DNA precipitated by the addition of 0.6 volumes propan-2-ol. The precipitate was collected by centrifugation at 8000rpm for 5 minutes at 4°C in a Sorvall GS-3 rotor, the pellet resuspended in 11ml TE and neutralised by the addition of 400ul 1M Tris base. 11g caesium chloride were then added and the resulting solution transferred to a 39ml Vti tube (Beckman). The tubes were filled to capacity with a 1.74g/ml aqueous solution of caesium chloride and ethidium bromide (Sigma) to a final concentration of 1mg/ml.

Following centrifugation at 45000rpm for 20-24 hours at 20 °C in a Beckman Vti50 rotor, supercoiled plasmid or cosmid DNA was collected by side puncture under longwave ultraviolet illumination.

The DNA solution was transferred to a Vti65 tube (Beckman) and centrifuged at 45000rpm for 20-24 hours at 20°C in a Beckman Vti65 rotor. Supercoiled DNA was isolated again by side puncture and ethidium bromide removed by extracting the preparation several times with water saturated butan-1-ol. After overnight dialysis against TE, DNA was precipitated by the addition of 0.1 volumes 2M sodium acetate and 2.5 volumes absolute ethanol. Following incubation for a minimum of 16 hours at -20°C, DNA was collected by centrifugation at 8000rpm for 15 minutes at 4°C in a Sorvall HB-4 rotor, the pellet washed thoroughly with 70% (v/v) ethanol, dried under vacuum, dissolved in a small volume of TE and stored at -20°C.

Section 2.7. Small scale preparation of cosmid and plasmid DNA.

Plasmid and cosmid DNA was prepared according to the method of Dr. E. Murray (NIMR, London). 10ml cultures of E.coli carrying plasmids or cosmids were grown at 37°C in L-broth or psi-broth, respectively, containing 100ug/ml ampicillin (Sigma) to late log phase. 1.5ml of this culture were centrifuged for 5 minutes at 4°C in an Eppendorf microfuge and the pellet resuspended in 200ul 50mM glucose, 25mM Tris-HCl, pH8.0, 10mM EDTA, pH8.0. 200ul of this suspension was gently mixed with 400ul of a freshly prepared solution of 0.2M sodium hydroxide, 1% (w/v) sodium dodecyl sulphate and incubated on ice for 5 minutes. Following the addition of 200ul ice-cold 3M potassium acetate, pH4.8, the mixture was incubated on ice for a further 5 minutes and centrifuged in an Eppendorf microfuge for 10 minutes at room temperature.

600ul of the resulting supernatant were extracted with an equal volume of phenol:chloroform:isoamylalcohol (50:50:1), pH7.0, residual phenol removed by ether extraction, and DNA precipitated by the addition of 0.6 volumes propan-2-ol. After centrifugation for 10 minutes at room temperature in an Eppendorf microfuge, the resulting DNA pellet was washed with 70% (v/v) propan-2-ol, dried under vacuum, resuspended gently in 50ul double distilled water, and stored at -20°C.

Section 2.8. Preparation of mammalian genomic DNA.

High molecular weight DNA was prepared by a modification of the procedure of Pellicer et al. (1978). Washed cell pellets, or nuclear pellets from RNA extraction procedure (section 2.10) were lysed at 10^7 nuclei/ml by incubation for 10-16 hours at 45°C in 100mM Tris-HCl, pH7.9, 10mM sodium chloride, 10mM EDTA, 0.5% (w/v) sodium dodecyl sulphate, 100ug/ml proteinase K (Sigma).

The lysate was shaken with an equal volume of phenol:chloroform:isoamylalcohol(50:50:1), pH7.0, and after centrifugation at 8000rpm for 15 minutes at 18°C in a Sorvall HB-4 rotor the aqueous phase was removed and mixed with 2 volumes 95% (v/v) ethanol. Precipitated nucleic acid was collected by centrifugation at 8000rpm for 15 minutes at 4°C in a Sorvall HB-4 rotor, the pellet dried under vacuum, resuspended in a minimum volume of 10mM Tris-HCl, pH7.4, 0.5mM EDTA, 10ug/ml RNase A (Sigma) and incubated at 37°C for 12-16 hours. The preparation was then dialysed at 4°C for 4-16 hours against TE and stored at 4°C.

Section 2.9. Preparation of total cellular RNA.

Total cellular RNA was prepared according to the method of Chirgwin et al. (1979) as modified by Dr. B.L.M. Hogan (NIMR, London). Cell monolayers were rinsed once with phosphate

buffered saline and lysed by the addition of 15ml 4M guanidinium thiocyanate (Fluka), 0.1M β-mercaptoethanol (Sigma), pH5.0, per 10^9 cells. The extract was gathered using a rubber policeman and stored at -70°C. Mouse tissues were carefully dissected, finely chopped in 10ml/g 4M guanidinium thiocyanate, 0.1M mercaptoethanol, pH5.0, and disrupted by 20 strokes in a glass dounce homogeniser on ice.

In both cases up to 30ml of extract were then layered over 5.5ml 5.7M caesium chloride, 0.1M EDTA, pH8.0, in a SW27 polyallomer tube (Beckman) and centrifuged at 27000rpm for 18 hours at 15°C in a Beckman SW27 rotor. Smaller volumes of extract, up to 11 ml, were layered onto 2.2ml 5.7M caesium chloride, 0.1M EDTA, pH8.0, in a SW41 polyallomer tube (Beckman) and centrifuged at 33000rpm for 24 hours at 15°C in a Beckman SW41 rotor.

After centrifugation the supernatant was aspirated, the tube carefully washed with sterile double distilled water, aspirated, and inverted to drain. The base of the tube was cut off using a sterile scalpel blade, the RNA pellet carefully rinsed, resuspended in a minimum volume of double distilled water, and transferred to a sterile, diethylpyrocarbonate (BDH)-treated, 1.5ml Eppendorf tube. Insoluble material was removed by centrifugation at room temperature for one minute in an Eppendorf microfuge, the supernatant transferred to a fresh tube, and nucleic acid precipitated by the addition of 7.5ul 4M sodium chloride per 100ul solution and 2.5 volumes 95% (v/v) ethanol. Following incubation at -20°C for a minimum

of 16 hours, RNA was collected by centrifugation for 5 minutes at 4^{o}C in an Eppendorf microfuge, the pellet washed twice with 70% (v/v) ethanol, dried under vacuum, resuspended in double distilled water and stored at -70^{o}C.

Section 2.10. Preparation of total cytoplasmic RNA.

Total cytoplasmic RNA was prepared by a modification of the procedure of Favalaro et al. (1980). Culture medium was removed from cell monolayers, the cells harvested in ice cold phosphate buffered saline (PBS) (137mM sodium chloride, 2.7mM potassium chloride, 8.1mM disodium hydrogen phosphate, 1.5mM potassium dihydrogen phosphate) by use of a rubber policeman, and pelleted by centrifugation at 2000rpm for 5 minutes at 4^{o}C in a Sorvall HB-4 rotor. The cell pellet was washed three times with ten volumes of ice cold PBS and resuspended at 10^{7}cells/ml in ice cold lysis buffer (140mM sodium chloride, 1.5mM magnesium chloride, 10mM Tris-HCl, pH8.6, 0.5% (v/v) Nonidet P40 (BDH), 10mM vanadyl ribonucleoside complex (BRL)) and vortexed briefly.

The lysate was underlayed with lysis buffer containing 1% (v/v) Nonidet P40, 30% (w/v) sucrose and nuclei pelleted by centrifugation at 8000rpm for 20 minutes at 4^{o}C in a Sorvall HB-4 rotor. The upper phase was removed to a fresh tube and mixed with an equal volume of PK buffer (200mM Tris-HCl, pH7.5, 25mM EDTA, pH8.0, 300mM sodium chloride, 2% (w/v)

sodium dodecyl sulphate, 200ug/ml proteinase K (Sigma)). Following incubation at 37°C for 30 minutes the solution was extracted with an equal volume of phenol: chloroform: isoamylalcohol (50:50:1), pH7.0, and centrifuged at 8000rpm for 15 minutes at 18°C in a Sorvall HB-4 rotor. The aqueous phase was removed to a clean tube and nucleic acid precipitated by the addition of 0.1 volumes 2M sodium acetate, 1.5 volumes propan-2-ol and incubation at -20°C for a minimum of 16 hours.

Nucleic acid was collected by centrifugation at 8000rpm for 45 minutes at -10°C in a Sorvall HB-4 rotor, the pellet washed twice with 70% (v/v) propan-2-ol, dried under vacuum, resuspended in sterile double distilled water and stored at -20°C.

Section 2.11. Selection of polyadenylated RNA on oligodT-cellulose.

Polyadenylated RNA was purified from total cellular RNA (section 2.9) or total cytoplasmic RNA (section 2.10) according to the method of Aviv and Leder (1972).

RNA samples diluted to 500ul with sterile double distilled water were mixed with an equal volume of 2x oligodT binding buffer (20mM Tris-HCl, pH7.5, 1M sodium chloride, 0.2% (w/v) sodium dodecyl sulphate) and incubated at 37°C for 2 minutes. The sample was cooled to room temperature and applied to a

0.5ml bed volume oligodT-cellulose (BRL) column equilibrated with 1x oligodT binding buffer. RNA was passed over the column six times, the column washed with five volumes of 1x oligodT binding buffer, five volumes of wash buffer (10mM Tris-HCl, pH7.5, 100mM sodium chloride, 0.1% (w/v) sodium dodecyl sulphate) and polyadenylated RNA eluted with three column volumes of sterile double distilled water.

Nucleic acid was precipitated by the addition of 0.1 volumes 2M sodium acetate, pH5.5, and two volumes propan-2-ol. Following incubation at -20^{o}C for 16 hours or more, the precipitate was collected by centrifugation at 8000rpm for 45 minutes at -10^{o}C in a Sorvall HB-4 rotor, the pellet washed with 70% (v/v) propan-2-ol, dried under vacuum, resuspended in sterile double distilled water and stored at -70^{o}C.

Section 2.12. Preparation of crude protein extracts.

E.coli cells bearing βgalactosidase-fusion vectors (section 2.1) were collected by centrifugation in a Sorvall RT6000B centrifuge at 2500rpm for 5 minutes at 4^{o}C. The pellet was washed once with ice cold PBS.

Mammalian cell monolayers (section 2.3) were washed once with ice cold PBS and collected in PBS using a rubber policeman. Cells were pelleted by centrifugation at 2000rpm for 5 minutes at 4^{o}C in a Sorvall RT6000B centrifuge and washed three times with ice cold PBS.

In both cases the cell pellet was resuspended in 300ul 1xSDS-PAGE loading buffer (Table 2.6) supplemented with 2mM phenylmethylsulphonyl fluoride (PMSF) (Sigma). Following sonication (two 10 second bursts using a Soniprobe 150 sonicator, 18um amplitude) on a salt-ice bath, cell lysates were stored at -20^{o}C. Samples were boiled for 5 minutes immediately prior to loading on an SDS polyacrylamide gel (section 2.15.4).

Section 2.13. Quantitation of nucleic acids.

2.13.1. Spectrophotometric assays.

Samples were diluted in TE and the absorbance at 260nm was measured. Concentrations were calculated assuming that an OD_{260} of 1.0 corresponds to 50ug/ml of double stranded DNA or 40ug/ml of RNA. The purity of nucleic acid preparations was assessed by calculation of the OD_{260}:OD_{280} ratio. Samples routinely gave a ratio of 1.8-2.0.

2.13.2. Radiometric assays.

Incorporation of ^{32}P and ^{35}S-labelled nucleotide into DNA or RNA molecules was measured by the method of acid precipitation (Maniatis et al. 1982).

An aliquot (0.01 reaction volumes) was applied to a Whatman GF/C glass fibre disc. Another aliquot was added to 100ul 0.5ug/ml sonicated salmon sperm DNA in 20mM EDTA, pH8.0. This

was added to 5ml ice cold 10% (v/v) trichloroacetic acid and incubated on ice for 15 minutes. The mixture was then applied to a Whatman GF/C glass fibre disc with suction, the disc washed with five volumes of ice cold 10% (w/v) trichloroacetic acid and one volume 96% (v/v) ethanol. Both filters were dried under infra-red illumination and counted in 5ml scintillation cocktail (Scintillator 299, United Technologies Packard) on either Cerenkovski (^{32}P) or ^{14}C (^{35}S) channels.

Section 2.14. Restriction endonuclease digestion.

Restriction enzymes used and their conditions of use are given in Table 2.5. Reactions were generally carried out using 50-100ug/ml DNA and 5 units enzyme/ug DNA. Incubations were performed at 37°C for 90 minutes (plasmid, cosmid, and bacteriophage lambda DNA) or 16 hours (mammalian genomic DNA). TaqI digests were performed at 65°C.

Conditions for partial digestion of high molecular weight mammalian genomic DNA were determined by the method of Maniatis et al. (1982) using two-fold serial enzyme dilutions. Restriction enzymes were purchased from Pharmacia PL, Boerhinger Mannheim, Anglian Biotechnology, and Northumbria Biologicals.

Table 2.5

Conditions for Restriction Enzyme Digestion of DNA.

	Restriction Enzyme Buffers		
	Low	Medium	High
Sodium chloride	0	50mM	100mM
Tris-HCl, pH7.5	10mM	10mM	50mM
Magnesium chloride	10mM	10mM	10mM
Dithiothreitol	1mM	1mM	1mM

Enzyme	Buffer
AccI	Low
AvaI	Medium
AvaII	Low
BamHI	High
BclI	Medium
BglI	High
BglII	High
ClaI	Low
DraI	Medium
EcoRI	Medium
HincII	Medium
HindIII	Medium
KpnI	Low
NdeI	High
NheI	Low
PstI	Medium

Table 2.5 cont.

PvuII	Medium
SacI	Low
SalI	High
SphI	High
StuI	Medium
TaqI	High
XbaI	High
XhoI	High
XhoII	Medium

Section 2.15. Electrophoretic techniques.

2.15.1. Neutral agarose gels.

Gels containing 0.4%-1.5% (w/v) agarose (Sigma) or low gelling temperature agarose (Sigma) were cast in TAE (Table 2.6) supplemented with 1ug/ml ethidium bromide (Sigma) according to the method of Favalaro et al. (1980). Immediately prior to loading, 0.1 volumes of DNA agarose gel loading buffer (Table 2.6) were added to all samples. Electrophoresis was carried out at 1.5V/cm for 4-16 hours. Fractionated DNA was viewed under longwave UV illumination.

2.15.2. Formaldehyde agarose gels.

RNA was fractionated on formaldehyde agarose gels (Lehrach et al. 1977). Gels containing 1% (w/v) agarose were prepared by boiling the agarose in 100ml 2xMOPS buffer (Table 2.6) and cooling to 65^{o}C. To this was added 84ml double distilled water prewarmed to 65^{o}C, the solution adjusted to 3% (v/v) formaldehyde (Fisons 37-41% (v/v)) and the gel poured.
Samples to be loaded were adjusted to 50% (v/v) formamide, 6% (v/v) formaldehyde in 1xMOPS buffer, incubated at 65^{o}C for 15 minutes and loaded immediately. Electrophoresis was performed in 1xMOPS buffer at 1.5V/cm for 15 hours. The size of specific RNA species was estimated by comparison with the rate of migration of a 0.3-9.5kb RNA ladder (BRL, cat.no. 520-5620).

Table 2.6

Electrophoretic Buffers.

DNA agarose gel loading buffer

0.25% (w/v) Bromophenol blue (BDH)

0.25% (w/v) Xylene cyanol (BDH)

25% (w/v) Ficoll400 (Sigma)

Stored at room temperature.

Formaldehyde gel loading buffer

50% (v/v) deionised formamide (Rose)

6% (v/v) formaldehyde 37-41% (v/v) (Fisons)

1xMOPS buffer

Stored at $4^{o}C$

SDS-PAGE loading buffer

4% (w/v) sodium dodecyl sulphate

20% (v/v) Glycerol

0.04% (w/v) Bromophenol blue

0.2M Dithiothreitol (Sigma)

0.16M Tris-HCl, pH6.8

Stored at $-20^{o}C$

1xTAE buffer

40mM Tris base

20mM acetic acid

1mM EDTA, pH8.0

Stored at room temperature

Table 2.6 cont.

1xMOPS buffer

20mM 3-[N-morpholino]propane-sulphonic acid (Sigma)

5mM sodium acetate, pH5.5

1mM EDTA, pH8.0

stored at 4°C

1xTBE buffer

90mM Tris base

90mM boric acid

2mM EDTA, pH8.0

stored at 4°C

1xSDS-PAGE buffer

0.19M glycine

0.1% (w/v) sodium dodecyl sulphate

25mM Tris-HCl, pH8.3

Acrylamide:Urea solution

5.7% (w/v) acrylamide (BDH)

0.3% (w/v) N,N' Methylene-bis-acrylamide (BDH)

7M urea (Sigma)

made to 800ml with distilled water. Deionised with 5g amberlite MB1 resin (BDH) for 30 minutes at 50°C, filtered, adjusted to 1 litre in 1x TBE and stored at 4°C.

2.15.3. Polyacrylamide thin gels.

20cm x 40cm x 0.4mm polyacrylamide gels were cast between clean, siliconised glass plates. To 50ml 6% (w/v) acrylamide:7M urea solution (Table 2.6) were added 0.75ml 10% (w/v) ammonium persulphate (BDH) and 75ul N,N,N'N' tetramethylethylenediamine (TEMED). The solutions were mixed, and the gel was poured immediately.
Products of sequencing reactions (section 2.22.3) were boiled for 3 minutes and loaded immediately. Electrophoresis was carried out in 1xTBE (Table 2.6) at 1500V constant voltage for 1-4 hours. Processing and autoradiography of gels were performed as described in section 2.21.

2.15.4. SDS polyacrylamide gel electrophoresis.

Proteins were electrophoresed in vertical SDS-polyacrylamide slab gels (Laemmli, 1970). The resolving gel (14cm) contained 5% (w/v) acrylamide, 0.13% (w/v) bisacrylamide, 0.1% (w/v) sodium dodecyl sulphate, 0.37M Tris-HCl, pH8.9, polymerisation being effected by the addition of 0.006 volumes 10% (w/v) ammonium persulphate and 0.001 volumes (v/v) TEMED. Following polymerisation the gel was overlaid with a stacking gel consisting of 5% (w/v) acrylamide, 0.13% (w/v) bisacrylamide, 0.1% (w/v) sodium dodecyl sulphate and 0.125M Tris-HCl, pH6.8, polymerisation being effected as before.
Samples in 1xSDS-PAGE loading buffer (Table 2.6) were boiled

for 5 minutes and loaded immediately. Electrophoresis was performed at 150V constant voltage for 3 hours in 1x SDS PAGE running buffer (Table 2.6).

Gels were stained by incubation in 25% (v/v) propan-2-ol, 10% (v/v) acetic acid, 0.05% (w/v) Coomassie blue for 16 hours with gentle agitation at room temperature. Following a further incubation for 5 hours in 10% (v/v) propan-2-ol, 10% (v/v) acetic acid, 0.0025% (w/v) Coomassie blue at room temperature with gentle agitation, gels were destained in 10% (v/v) acetic acid for 2 hours at room temperature.

Section 2.16. Molecular cloning techniques.

2.16.1. Purification of DNA fragments from agarose gels.

DNA fragments for subcloning were prepared according to the method of Vogelstein and Gillespie (1979) as modified by Dr. M. Lee (Imperial Cancer Research Fund).

Finely ground glass particles were obtained by thoroughly grinding five glass scintillation vials in pestle and mortar. The glass particles were stirred in three litres of double distilled water and the suspension allowed to stand for 40 minutes. One litre of the suspension was carefully decanted into a clean beaker and the glass particles collected by centrifugation at 10000rpm for 10 minutes at 20^{o}C in a Sorvall GS-3 rotor. The pellet was resuspended in a small volume of

double distilled water, transferred to a preweighed beaker, dried under vacuum, resuspended at 100ug/ml in double distilled water, autoclaved and stored at room temperature.

An agarose gel slice containing the restriction fragment of interest was added to a preweighed 1.5ml Eppendorf tube and dissolved in two volumes of a saturated solution of sodium iodide (BDH). Ground glass suspension was added to a final concentration of 10mg/ug DNA and DNA allowed to adsorb by incubation for 12-16 hours with gentle agitation at room temperature. Glass particles were pelleted by centrifugation for one minute at room temperature in an Eppendorf microfuge and washed with 70% (w/v) sodium iodide for 20 minutes at room temperature with gentle agitation. Following centrifugation for one minute in an Eppendorf microfuge the pellet was washed once with 50% (v/v) ethanol, resuspended in 400ul 20mM Tris-HCl, pH7.5, 200mM sodium chloride, 2mM EDTA, pH8.0 and incubated at 37^{o}C for 3-4 hours with constant agitation.

Glass particles were pelleted by centrifugation in an Eppendorf microfuge for 10 minutes at room temperature, the supernatant removed to a clean 1.5ml Eppendorf tube and DNA precipitated by addition of 0.1 volumes 2M sodium acetate, pH 5.5 and 2.5 volumes of absolute ethanol. After incubation at -20^{o}C for 16 hours or more, precipitated DNA was collected by centrifugation for 15 minutes at 4^{o}C in an Eppendorf microfuge. The pellet was washed with 70% (v/v) ethanol, dried under vacuum, resuspended in a minimum volume of double distilled water and stored at -20^{o}C.

2.16.2. Purification of partially digested genomic DNA restriction fragments.

Size selection of the products of partial restriction digests of mouse genomic DNA (section 2.14) were performed by centrifugation through a 10-40% (w/v) sucrose density gradient as described by Maniatis et al. (1982).

A 10-40% (w/v) sucrose gradient was poured in a SW27 tube (Beckman) using sucrose solutions in a buffer of 1M sodium chloride, 20mM Tris-HCl, pH8.0, 5mM EDTA, pH8.0. 500ug of partially digested genomic DNA were loaded onto the gradient and centrifugation performed in a SW27 rotor at 26000rpm for 24 hours at 15^{o}C.

After centrifugation, a 21 guage needle was inserted through the bottom of the tube and 0.5ml fractions collected. To 15ul of every third fraction was added 4ul DNA agarose gel loading buffer (Table 2.6) and electrophoresis performed as described in section 2.15.1. Appropriately sized fractions were pooled and dialysed against TE for 16 hours at 4^{o}C. The volume of the resulting solution was reduced by extraction with butan-1-ol and nucleic acid precipitated by the addition of 0.1 volumes 2M sodium acetate, pH5.5, and 2.5 volumes of 96% (v/v) ethanol.

Nucleic acid was collected by centrifugation, the pellet washed with 70% (v/v) ethanol, dried briefly under vacuum, gently resuspended in a small volume of TE and stored at 4^{o}C.

2.16.3. Dephosphorylation of DNA fragments.

5′ phosphate groups were removed from plasmid vector DNA fragments by use of calf intestinal alkaline phosphatase (CAP) essentially as described by Maniatis et al. (1982).

To a maximum of 1ug of purified DNA fragment (section 2.16.1) were added 5ul 10x CAP buffer (0.5M Tris-HCl, pH9.0, 10mM magnesium chloride, 1mM zinc chloride, 10mM spermidine (Sigma)), double distilled water to a total of 50ul and 0.0025 units of calf intestinal alkaline phosphatase (Boerhinger Mannheim). Dephosphorylation of recessed 5′ termini or blunt ended fragments was performed by incubation at 37°C for 15 minutes and 56°C for 15 minutes. A second aliquot of enzyme was added and the incubations repeated. Dephosphorylation of protruding 5′ termini was performed by incubation at 37°C for 30 minutes, the addition of a second aliquot of enzyme, and incubation at 37°C for a further 30 minutes.

The reaction mix was made up to 200ul with TE, pH7.5, extracted with phenol:chloroform:isoamylalcohol (50:50:1), pH7.0, and the organic phase back extracted with TE, pH7.5. Following ether extraction DNA was precipitated by the addition of 0.1 volumes 2M sodium acetate, pH5.5, 2.5 volumes of 96% (v/v) ethanol and incubation at -20°C for a minimum of 16 hours. DNA was pelleted by centrifugation for 15 minutes at 4°C in an Eppendorf microfuge, the pellet washed with 70% (v/v) ethanol, dried under vacuum, resuspended in a small volume of double distilled water and stored at -20°C.

2.16.4. Conversion of cohesive termini to blunt ended DNA fragments.

The conversion of DNA fragments with protruding 5′ or 3′ termini to blunt ended fragments was performed by the method of Maniatis et al. (1982).

Protruding 3′ termini were removed from up to 1ug of purified restriction fragment (section 2.16.1) by incubation of DNA in 33mM Tris-acetate, pH7.9, 66mM potassium acetate, 10mM magnesium acetate, 0.5mM dithiothreitol (Sigma), 0.1uM dCTP, dTTP, dATP, dGTP, 1mg/ml bovine serum albumin and 2.5 units T4 DNA polymerase (Pharmacia), in a total reaction volume of 20ul, for 5 minutes at 37^{o}C.

Fragments with recessed 3′ termini were filled by incubating up to 1ug of restriction fragment in 50mM Tris-HCl, pH7.2, 10mM magnesium sulphate, 0.1mM dithiothreitol, 50ug/ml bovine serum albumin, 0.1uM dCTP, dTTP, dATP, dGTP, and 2 units Klenow fragment of E.coli DNA polymerase I (Pharmacia), in a total reaction volume of 25ul, for 30 minutes at 25^{o}C.

In both cases, the volume was made to 200ul with TE, pH7.5, extracted once with neutral phenol:chloroform:isoamylalcohol (50:50:1) and the organic phase back extracted with TE, pH7.5. The pooled aqueous phases were extracted once with water saturated ether and DNA precipitated by the addition of 0.1 volumes 2M sodium acetate, pH5.5, and 2.5 volumes 96% (v/v) ethanol. Following incubation at -20^{o}C for a minimum of 16 hours DNA was collected by centrifugation at 4^{o}C for 15

minutes in an Eppendorf microfuge, the pellet washed with 70% (v/v) ethanol, dried under vacuum, resuspended in a small volume of sterile double distilled water, and stored at -20°C.

2.16.5. Partial filling-in of cohesive termini.

Genomic DNA fragments with cohesive Sau3a termini were partially filled-in essentially as described by Zabarovsky and Allikmets (1986). 5ug of Sau3a partially digested mouse genomic DNA (section 2.14) were incubated in 50mM Tris-HCl, pH 8.0, 7mM magnesium chloride, 1mM dithiothreitol, 330uM dGTP, 330uM dATP, 22.5 units Klenow fragment of E.coli DNA polymerase I (Pharmacia) in a final reaction volume of 30ul. Following incubation at 25°C for 20 minutes and 68°C for 10 minutes the solution was extracted with phenol:chloroform: isoamylalcohol (50:50:1), pH7.0, and DNA precipitated by the addition of 0.1 volumes 2.5M sodium acetate, pH5.5, and 2.5 volumes 96% (v/v) ethanol. After incubation at -20°C for 16 hours precipitated DNA was collected by centrifugation, the pellet washed with 70% (v/v) ethanol, dried briefly under vacuum, resuspended in 5ul TE and stored at 4°C.

2.16.6. Ligation of DNA fragments.

A 3:1 molar ratio of insert DNA:dephosphorylated vector DNA was added to 1ul 660mM Tris-HCl, pH7.6, 66mM magnesium chloride, 100mM dithiothreitol, 1ul 100mM ATP (Pharmacia), and

2 units T4 DNA ligase (BRL) in a total reaction volume of 10ul. Cohesive-end ligations were carried out at 14°C for 6 hours followed by incubation at 37°C for 60 minutes. Blunt-end ligations were performed at 4°C for 20 hours, 14°C for 1 hour, and 37°C for 1 hour. In both cases the reaction was terminated by the addition of 1ul 500mM EDTA, pH8.0 and stored at 4°C.

2.16.7. Transformation of E.coli K12 strains.

E.coli strains BMH71-18 and DH5 were made competent for transformation with plasmid DNA according to the method of Scott, (1982).

A single bacterial colony was used to inoculate 5ml psi-broth (Table 2.2) in a 500ml flask and incubated with vigorous aeration at 37°C until OD_{550}=0.30. This culture was then added to 100ml psi-broth prewarmed to 37°C in a 500ml flask and incubated with vigorous aeration at 37°C until OD_{550}=0.48.

The culture was rapidly cooled in iced water and cells collected by centrifugation at 2500rpm for 5 minutes at 4°C in a Sorvall RT6000B centrifuge. The cell pellet was resuspended in 30ml 100mM rubidium chloride (Sigma), 50mM manganese chloride (Sigma), 30mM potassium acetate, 10mM calcium chloride, 15% (v/v) glycerol. After incubation on ice for 5 minutes (BMH71-18) or 15 minutes (DH5) the cells were collected by centrifugation at 2500rpm for 5 minutes at 4°C in a Sorvall RT6000B centrifuge. The pellet was resuspended

gently in 4ml 10mM 3-(N-morpholino)propanesulphonic acid, pH7.0, 10mM rubidium chloride, 75mM calcium chloride, 15% (v/v) glycerol, incubated on ice for 15 minutes (BMH71-18) or 20 minutes (DH5), dispensed into 200ul aliquots and stored at -70^{o}C.

For transformation with plasmid DNA, competent cells were thawed and incubated on ice for 10 minutes. No more than 100ng DNA or 5ul of ligation mix (section 2.16.6) were then added per 200ul competent cells, the mixture incubated on ice for 20 minutes, 37^{o}C for 2 minutes, followed by a further incubation on ice for 2 minutes. 4 volumes of psi-broth were added and, following incubation for 60 minutes at 37^{o}C with gentle shaking, cells were collected by centrifugation at 2500rpm for 5 minutes in a Sorvall RT6000B centrifuge. The pellet was gently resuspended in 200-500ul psi-broth and 100ul aliquots were plated onto L-agar plates supplemented with 100ug/ml ampicillin (Sigma) and incubated at 37^{o}C for 12-18 hours.

To prepare competent E.coli JM103, a single colony was picked from a minimal agar plate (Table 2.2) into 25ml L-broth and incubated at 37^{o}C with vigorous aeration until OD_{550}=0.4-0.6. Cells were chilled rapidly on ice and collected by centrifugation at 2500rpm for 5 minutes at 4^{o}C. The pellet was resuspended in 0.5 volumes ice cold 50mM calcium chloride and incubated on ice for 20 minutes. Cells were pelleted as before, resuspended in 0.1 volumes 50mM calcium chloride and stored at 4^{o}C.

5ul ligation mix (section 2.16.6) or up to 1ng M13 or pBSM13^{+}

DNA (Table 2.3) was added to 300ul of cells, the mixture incubated on ice for 40 minutes, and heat shocked at 45°C for 2 minutes. When transforming with M13 DNA, 3ml top agarose (Table 2.2) were then added along with 20ul 25mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) (Sigma), 20ul 25mg/ml isopropyl β-D thiogalactoside (IPTG) (Sigma), and the mixture poured onto L-agar plates.

When transforming with pBSM13$^+$ plasmid DNA, 1ml L-broth was added and the mixture incubated at 37°C for 50-60 minutes with gentle shaking. Cells were collected by centrifugation at 2500rpm for 5 minutes at room temperature in a Sorvall RT6000B centrifuge, the pellet resuspended in 300ul L-broth supplemented with 100ug/ml ampicillin and plated onto L-agar plates containing 100ug/ml ampicillin, 40ug/ml X-gal, and 5mM IPTG.

Following incubation at 37°C for 12-18 hours plates were incubated at 4°C for 2-4 hours in order to enhance the blue/white colour contrast.

2.16.8. In vitro packaging of bacteriophage DNA.

Products of ligation reactions involving bacteriophage lambda DNA fragments (section 2.16.6) were packaged in vitro into bacteriophage particles by the use of Gigapack Gold packaging extracts (Stratagene). 4ul of ligation mix were added to 10ul freeze/thaw lysate extract on ice in a 1.5ml Eppendorf tube. To this was added 15ul sonicated extract, the contents mixed

gently and incubated at 22°C for 2 hours. The mixture was then made to 500ul with SM buffer (Table 2.2), a drop of chloroform added, and stored at 4°C.

Section 2.17. Radiolabelling of nucleic acids.

2.17.1. Radiolabelling of DNA restriction endonuclease fragments.

Fragment preparation.

DNA was radiolabelled <u>in vitro</u> with [α^{32}P] dCTP by use of a modified version of the "oligolabelling" method described by Feinberg and Vogelstein (1984).

2-10ug of plasmid, bacteriophage, or cosmid DNA was digested with the appropriate restriction endonucleases (section 2.14) and the resulting DNA fragments separated by electrophoresis through 1-2% (w/v) low gelling temperature agarose gels as described in section 2.15.1).

A gel band containing the restriction fragment of interest was carefully excised and placed in a preweighed 1.5ml Eppendorf tube to which three volumes of double distilled water were added. The gel fragment was broken up, the gel suspension placed in a boiling water bath for 7 minutes, quenched on ice, and stored at -20°C.

Labelling Reaction.

The fragment preparation was placed in a boiling water bath for 3 minutes and incubated at 37^{o}C for 10 minutes. Up to 34ul of this suspension was then added to 10ul oligolabelling buffer (1M HEPES, pH 6.6, 0.25M Tris-HCl, pH 8.0, 25mM magnesium chloride, 50mM mercaptoethanol, 0.1mM dATP, dGTP, dTTP, 270OD_{260}/ml hexamer DNA random primers (Pharmacia)), 2ul 10mg/ml bovine serum albumin (Sigma), 35uCi [α^{32}P] dCTP (Amersham International, specific activity 3000Ci/mmol) in a total reaction volume of 50ul. The reaction was initiated by the addition of 5 units Klenow fragment E.coli DNA polymerase I (Pharmacia) and incubated at room temperature for 6 hours (30-100ng DNA template) or 16 hours (10-30ng DNA template). Labelled DNA was purified by passage over a 2.5ml column of Sephadex G50 (Pharmacia) equilibrated in TE. The excluded peak of radioactivity was pooled and stored at -20^{o}C.

Routinely 60-70% of radioactivity was incorporated into the excluded peak, as measured by TCA precipitation (section 2.13.2) giving specific activities of 5x10^{7} cpm/ug to 1x10^{8} cpm/ug. By using up to 50uCi [α^{32}P] dCTP specific activities up to 10^{9} cpm/ug were obtained.

2.17.2. In vitro synthesis of high specific activity RNA probes.

Template preparation.

DNA fragments were subcloned into plasmid vectors pSP64, pSP65, or pBSM13$^+$ as described in section 2.16. Following restriction endonuclease digestion (section 2.14), proteinase K was added to a final concentration of 200ug/ml. After incubation at 37^{o}C for 30 minutes, the mixture was extracted with phenol: chloroform: isoamylalcohol (50:50:1), pH7.0, and DNA precipitated by the addition of 0.1 volumes 2M sodium acetate, pH5.5, 2.5 volumes 96% (v/v) ethanol and incubation at -20^{o}C for 16 hours or more. DNA was pelleted by centrifugation in an Eppendorf microfuge for 10 minutes at 4^{o}C, the pellet washed with 70% (v/v) ethanol, dried briefly under vacuum, resuspended in a small volume of RNase-free water and stored at -20^{o}C.

Labelling reaction.

Routinely 1ug of DNA template was used in each labelling reaction. To this DNA template was added 5ul 5x transcription buffer (200mM Tris-HCl, pH8.0, 40mM magnesium chloride, 10mM spermidine, 250mM sodium chloride), 1ul 10mM ATP, 1ul 10mM UTP, 1ul 10mM GTP (Stratagene), 1ul 0.75M dithiothreithol, 20 units RNasin (Promega Biotec), 100uCi [α^{32}P] CTP (Amersham International, specific activity 400-800Ci/mmol), or 100uCi [α^{35}S] UTP (NEN 1000-1500Ci/mmol), 10 units T3 or T7

polymerase (Stratagene), as appropriate, in a final reaction volume of 25ul.

Following incubation at 37^{o}C for 30 minutes 1ul vanadyl ribonucleoside complex (BRL) and RNase-free DNaseI (BRL) were added to a final concentration of 0.5ug/ul and the mixture incubated at 37^{o}C for a further 10 minutes.

The reaction mix was then loaded onto a G50 sephadex column and eluted with RNase-free TE, 10mM dithiothreitol, 0.1% (w/v) sodium dodecyl sulphate. The first peak was pooled and nucleic acid precipitated by the addition of 0.4 volumes 5M ammonium acetate, 2 volumes propan-2-ol, and incubated for 10 minutes at room temperature. Following centrifugation in an Eppendorf microfuge for 15 minutes at room temperature, the pellet was dried briefly under vacuum, resuspended in 10mM dithiothreitol, and stored at -20^{o}C.

Hydrolysis of RNA probes.

Hydrolysis of RNA probes for use in RNA _in situ_ hybridisation experiments (section 2.20.5) was performed according to the method of Cox _et al_. (1984).

The probe preparation was adjusted to 40mM sodium hydrogen carbonate, 60mM sodium carbonate, pH 10.2, and incubated at 60°C for t minutes:

$$t = \frac{L_0 - L_f}{K.L_0.L_f}$$

t = time (minutes)

L_0= original probe length (kb)

L_f= final probe length (kb)

K = constant = 0.11 $kb^{-1}min^{-1}$

An average size of 100-150 nucleotides was generated. Hydrolysis was terminated by addition of 0.05 volumes 2M sodium acetate, pH5.5, 0.05 volumes 10% (v/v) acetic acid, and RNA precipitated with 2 volumes ethanol.

Following overnight incubation at -20°C, RNA was collected by centrifugation for 15 minutes at room temperature, the pellet dried briefly, resuspended in 10mM dithiothreitol at a concentration of 3ng/ul/kb length of original probe and stored at -20°C.

Section 2.18. Filter preparation for nucleic acid hybridisations.

2.18.1. Filter preparation for RNA gel transfer hybridisation.

Following electrophoresis (section 2.15.2), excess gel was carefully removed with a scalpel blade and excess formaldehyde removed by rinsing several times in distilled water. The gel was incubated successively in 0.05M sodium hydroxide for 30 minutes, 0.1M Tris-HCl, pH7.5, for 30 minutes, and 1xSSC (150mM sodium chloride, 15mM tri-sodium citrate, pH 7.0) for 20 minutes, all with gentle agitation at room temperature.
The gel was then carefully placed on two sheets of Whatman 3MM paper soaked in, and in contact with, a reservoir of 1xSSC, care being taken to avoid trapping air bubbles. A piece of Genescreen (NEN) cut to the size of the gel and soaked in 1xSSC for 20 minutes was placed on the surface of the gel avoiding trapping air bubbles. Five pieces of dry Whatman 3MM paper were placed on top of the filter followed by several layers of paper tissue, a glass plate, and a 1kg weight. Transfer was allowed to proceed for 12-18 hours at $4^{o}C$.
The filter was then removed, washed thoroughly in 1xSSC, air dried, and baked at $80^{o}C$ for 4 hours.

2.18.2. Filter preparation for DNA gel transfer hybridisation.

DNA restriction fragments fractionated by agarose gel electrophoresis (section 2.15.1) were transferred from agarose gels to nitrocellulose membranes by the procedure of Southern (1975) as modified by Wahl et al. (1979).

Excess gel was removed with a scalpel blade and the gel incubated in 0.25M HCl for 15 minutes at room temperature with gentle agitation. After incubation in 0.2M sodium hydroxide, 0.6M sodium chloride for 30 minutes at room temperature, the gel was neutralised with 0.5M Tris-HCl, pH7.6, 1.5M sodium chloride for 30 minutes at room temperature.

The gel was then placed on two sheets of Whatman 3MM paper soaked in, and in contact with, a reservoir of 2xSSC, care being taken to avoid trapping air bubbles. A piece of nitrocellulose membrane (Schleicher & Schuell) cut to the size of the gel and soaked for 20 minutes in 2xSSC was then placed over the gel avoiding trapping air bubbles. Five pieces of Whatman 3MM paper were placed on top of the nitrocellulose followed by several thicknesses of tissue paper. Close contact between the components of the transfer system was maintained by the application of a 1kg weight.

Transfer was allowed to proceed for 20 minutes at room temperature for plasmid, cosmid, or bacteriophage lambda DNA fragments and for 12-18 hours at $4^{o}C$ for mammalian genomic DNA fragments. Up to four filters were prepared from a single gel

of plasmid, cosmid, or bacteriophage lambda DNA fragments. Following transfer the filter was washed thoroughly in 2x SSC, air dried and baked at 80°C for 2 hours.

2.18.3. Filter preparation for hybridisation screening of bacterial colonies.

E.coli colonies bearing plasmids or cosmids were prepared for hybridisation screening by a modification of the method of Grunstein and Hogness (1975).

Cosmid libraries were plated on 21cm x 21cm L-agar plates (Table 2.2) containing 100ug/ml ampicillin overlaid with a sheet of nitrocellulose (Schleicher & Schuell) at a density of 200000 colonies per plate and grown at 37°C for 12-18 hours. The nitrocellulose sheet was then removed and placed colony side uppermost on three sheets of Whatman 3MM paper. Duplicate filters were prepared by placing a fresh sheet of nitrocellulose over the colony-containing sheet followed by three sheets of Whatman 3MM paper and a glass plate. Heavy pressure was applied to ensure good transfer. Two duplicate filters were prepared from each master plate and orientation marks applied. The replica filters were then placed on L-agar plates supplemented with 100ug/ml ampicillin and grown at 37°C for 6-10 hours.

Transformation products from plasmid subcloning experiments (section 2.16), or cosmid clones picked from a primary screen of a cosmid library, were plated on L-agar plates containing

100ug/ml ampicillin and grown at 37°C for 12-18 hours. Following incubation at 4°C for 30-60 minutes the plate was overlaid with a nitrocellulose disc (Schleicher & Schuell). After orientating the filter and plate, the filter was removed and placed colony side uppermost on a sheet of Whatman 3MM paper. The plates were reincubated at 37°C for 6-10 hours and stored at 4°C.

Filters prepared by either method were incubated in 0.5M sodium hydroxide for 5 minutes with colony side uppermost, care being taken to soak only the underside of the filter. Following neutralisation in 0.1M Tris-HCl, pH7.5, for 5 minutes at room temperature filters were incubated in 1.5M sodium chloride, 0.5M Tris-HCl, pH7.5, for 10 minutes. Bacterial debris was removed by scrubbing the filters thoroughly in 2xSSC, 0.1% (w/v) sodium dodecyl sulphate. After incubation in 2xSSC for 10 minutes at room temperature filters were air dried and baked for 2 hours at 80°C.

2.18.4. Filter preparation for screening of bacteriophage lambda plaques.

An appropriate number of plaque forming units were plated onto L-agar or NZCYM plates as described in section 2.2 and incubated at 37°C for 12-18 hours.

Following incubation at 4°C for 30-60 minutes, plates were overlaid with a Biodyne A nylon membrane (Pall Corporation). Orientation marks were applied to the filter and plate, the

filter removed and placed plaque side uppermost on a sheet of Whatman 3MM paper. A second membrane was then applied to the plate and the process repeated.

The filters were processed by incubation in 1.5M sodium chloride, 0.5M sodium hydroxide for 5 minutes at room temperature followed by neutralisation in 3M sodium acetate, pH5.5, for 5 minutes. The filters were washed briefly in 2xSSC, air dried and baked for 1 hour at 80^{o}C.

Section 2.19. Preparation of embryo and tissue sections for RNA <u>in</u> <u>situ</u> hybridisation.

Fixation and embedding.

Mouse embryos or tissues were carefully dissected and incubated in 5-10ml 4% (w/v) paraformaldehyde (Sigma) at 4^{o}C for 12-16 hours. This solution was replaced with 5ml phosphate buffered saline, then 5ml 0.85% (w/v) saline, each for 30-60 minutes at 4^{o}C.

The solution was replaced with 3ml 0.85% (w/v) saline over which was layered an equal volume of 100% ethanol. The two solutions were gently but thoroughly mixed and left at room temperature for 15-30 minutes. The embryos or tissues were then incubated successively in ethanol:saline (1:1), 70% (v/v) ethanol, and 70% (v/v) ethanol each for 15-30 minutes at room temperature. Dehydration was completed by incubation in 85% (v/v) ethanol, 95% (v/v) ethanol, 100% ethanol, and 100%

ethanol each for 30-60 minutes at room temperature.

Embryos and tissues were subsequently incubated in two changes of toluene, each for 30-60 minutes at room temperature, followed by three incubations, each for 20-40 minutes, in fibrowax pastillated paraffin wax (Raymond Lamb) at 60^{o}C prior to embedding in paraffin wax. Once set, wax blocks were stored at 4^{o}C until required.

Pretreatment of slides.

Microscope slides were washed briefly in 10% (v/v) hydrochloric acid in 70% (v/v) ethanol, distilled water, 95% (v/v) ethanol, then dried at 150^{o}C for 5 minutes. Slides were then immersed in 0.4% (w/v) gelatin, 0.4mg/ml potassium chromium sulphate (Aldrich), allowed to drain vertically, baked at 150^{o}C for 15 minutes, and stored at room temperature until required.

Sectioning.

Wax blocks containing embryos or tissues were trimmed, fixed to a wooden block with wax, and the block set on a microtome. Ribbons of 6um sections were cut and placed on the surface of a 50^{o}C waterbath for 1-2 minutes. The ribbons were then lifted onto a treated microscope slide, dried at 50^{o}C for 30 minutes and stored in desiccant at room temperature for 12-16 hours. Sections were then stored at 4^{o}C with desiccant until required.

Pretreatment of sections.

Slides were dewaxed by two incubations in xylene for 10 minutes each time and xylene removed with 100% ethanol. Sections were then rehydrated by quickly immersing slides in 100%, 100%, 95% (v/v), 85% (v/v), 70% (v/v), 50% (v/v), and 30% (v/v) ethanol prior to incubation in 0.85% (w/v) saline for 5 minutes.

Following a 5 minute incubation in phosphate buffered saline (PBS), one 20 minute incubation in 4% (w/v) paraformaldehyde, and two 5 minute washes with PBS, the slides were drained, placed on a horizontal surface, and the sections overlaid with 20ug/ml proteinase K (Sigma) in 50mM Tris-HCl, 5mM EDTA, pH7.5. After incubation at room temperature for 5 minutes the slides were washed again with PBS and refixed in 4% (w/v) paraformaldehyde for 5 minutes.

Slides were quickly dipped in distilled water prior to immersion in 250ml 0.1M triethanolamine HCl, pH8.0, (Sigma) containing 0.625ml acetic anhydride (Sigma). After 10 minutes incubation the slides were washed with PBS, then 0.85% (w/v) saline, each for 5 minutes, dehydrated quickly by putting through 30% (v/v), 50% (v/v), 70% (v/v), 85% (v/v), 95% (v/v), 100%, and 100% ethanol, air dried and used immediately for hybridisation (section 2.20.5).

Section 2.20. Hybridisation techniques.

2.20.1. Hybridisation of radioactive DNA probes to RNA gel transfer filters.

RNA filters, prepared as in section 2.18.1, were incubated in 10-20ml hybridisation buffer (50% (v/v) deionised formamide (Rose Chemicals), 5xSSC, 50mM sodium phosphate, pH6.4, 0.1% (w/v) sodium dodecyl sulphate, 0.2mg/ml ficoll400 (Sigma), 0.2mg/ml polyvinylpyrrolidone (Sigma), 0.2mg/ml bovine serum albumin, 0.5mg/ml sonicated salmon sperm DNA (Sigma) in sealed plastic at 42^{o}C for 24 hours with gentle agitation.
The radioactive DNA probe (section 2.17.1) was denatured by incubation in a boiling waterbath for 10 minutes followed by rapid cooling on ice. The hybridisation solution was removed and replaced with fresh buffer containing denatured DNA probe at a concentration of approximately 10^{6}cpm/ml. Following incubation for 16-20 hours at 42^{o}C with constant agitation the filter was washed three times in several volumes of 2xSSC, 0.1% (w/v) sodium dodecyl sulphate at 55^{o}C for 30 minutes with gentle agitation. After three 30 minute washes with 0.2xSSC, 0.1% (w/v) sodium dodecyl sulphate at 55^{o}C filters were air dried and autoradiography performed as described in section 2.21.2.

2.20.2. Hybridisation of radioactive RNA probes to RNA gel transfer filters.

These were performed as for DNA probes (section 2.20.1) but the hybridisation buffer consisted of 50% (v/v) deionised formamide, 5xSSC, 50mM Tris-HCl, pH7.5, 0.1% (w/v) sodium pyrophosphate, 1% (w/v) sodium dodecyl sulphate, 0.2% (w/v) polyvinylpyrrolidone, 0.2% (w/v) ficoll400, 5mM EDTA, pH8.0, 150ug/ml sonicated salmon sperm DNA. Prehybridisation, hybridisation and washes were all performed at 65^{o}C.

2.20.3. Hybridisation of radioactive DNA probes to DNA gel transfer filters.

DNA filters prepared in section 2.18 were incubated in sealed plastic with 10-20ml hybridisation buffer (3xSSC, 0.1% (w/v) sodium dodecyl sulphate, 2mg/ml ficoll400, 2mg/ml polyvinylpyrrolidone, 2mg/ml bovine serum albumin, 50ug/ml sonicated salmon sperm DNA) at 65^{o}C with gentle agitation for 6-24 hours.

The radioactive DNA probe containing 1ug/ml plasmid vector DNA as competitor was denatured by incubation for 10 minutes in a boiling water bath followed by rapid cooling on ice. The hybridisation solution was replaced with fresh buffer containing the denatured DNA probe and plasmid competitor at a concentration of approximately 10^{6}cpm/ml.

Hybridisation was allowed to proceed for 16-20 hours at 65^{o}C

with constant agitation. The filters were subsequently washed three times for 20 minutes each time with several volumes of 2xSSC, 0.1% (w/v) sodium dodecyl sulphate and three times for 20 minutes with 0.2xSSC, 0.1% (w/v) sodium dodecyl sulphate, all washes being performed at 65°C. The filters were air dried and autoradiography carried out as described in section 2.21.2.

2.20.4. Hybridisation of radioactive RNA probes to DNA gel transfer filters.

These were performed as described in section 2.20.3 except that the hybridisation buffer used was the same as that in section 2.20.2. Prehybridisation and hybridisations were carried out at 42°C and filters washed at 65°C.

2.20.5. Hybridisation of RNA probes to embryo and adult tissue sections.

Hybridisation of ^{35}S-labelled RNA probes to embryo and tissue sections were performed essentially as described by Wilkinson et al. (1987a).

Hydrolysed probe (section 2.17.2) was diluted ten-fold in 50% (v/v) formamide, 0.3M sodium chloride, 20mM Tris-HCl, pH8.0, 5mM EDTA, 10mM sodium phosphate, pH8.0, 10% (w/v) dextran sulphate, 100mg/ml Ficoll, 100mg/ml polyvinylpyrrolidone, 100mg/ml bovine serum albumin, 0.5mg/ml yeast RNA, heated at

80°C for 2 minutes, and then degassed at room temperature for 5-10 minutes.

7-10ul of hybridisation solution was overlaid onto each section and a siliconised glass coverslip placed on top ensuring to trap no air bubbles. Slides were placed horizontally in a slide box containing tissue saturated with 50% (v/v) formamide, 5xSSC, the box sealed, and immersed in a 50°C waterbath for 16-18 hours.

The slides were then removed and incubated in 250ml degassed 5xSSC, 10mM dithiothreitol at 50°C for 15-30 minutes to allow the coverslips to dislodge. Following a 20 minute incubation in 50% (v/v) formamide, 0.1M dithiothreitol at 65°C, the slides were washed three times in 0.5M sodium chloride, 10mM Tris-HCl, pH7,5, 5mM EDTA at 37°C for 10 minutes prior to incubation for 30 minutes at 37°C in the same solution supplemented with 20ug/ml RNase A.

Following three further washes in the salt, Tris, EDTA solution at 37°C, the sections were incubated in 50% (v/v) formamide, 2xSSC, 0.1M dithiothreitol at 65°C for 20 minutes, 2xSSC for 15 minutes, 0.1xSSC for 15 minutes, both at room temperature, then dehydrated quickly by putting through 30% (v/v), 60% (v/v), 80% (v/v), and 95% (v/v) ethanol containing 0.3M ammonium acetate. After two rinses with 100% ethanol the slides were air dried and autoradiography performed as described in section 2.21.3.

Section 2.21. Autoradiography.

2.21.1. Polyacrylamide thin gels.

Gels were fixed with 5% (v/v) acetic acid, 5% (v/v) methanol for 15 minutes with gentle agitation. The gel was then transferred to Whatman 3MM paper, overlaid with Saran Wrap, dried under vacuum at 80^{o}C for 60 minutes, the Saran Wrap removed, and the gel exposed directly to Fuji X-ray film at room temperature.

2.21.2. Nitrocellulose and nylon membranes.

Air dried filters were attached to a sheet of Whatman 3MM paper and overlaid with Saran Wrap. Biodyne A nylon membranes were sealed whilst damp between two sheets of Saran Wrap. Autoradiography was performed at -70^{o}C using Fuji X-ray film or Kodak XAR-5 film with Hi-Speed-X intensifying screens as described by Laskey and Mills (1977).

2.21.3. Embryo and tissue sections.

Localisation of ^{35}S RNA probes on embryo and tissue sections was determined by use of Ilford K5 nuclear track emulsion. A Kodak Wratten II safelight was used throughout. Emulsion shreds were melted at 43^{o}C and 6ml molten emulsion was gently but thoroughly mixed with an equal volume of a 2% (v/v)

glycerol solution prewarmed to 43°C. A clean slide was used to ensure the emulsion gave a smooth coating and each experimental slide was dipped, allowed to drain vertically and placed horizontally in an aerated plastic box.

The box was then wrapped in foil and left at room temperature for 2 hours. A sachet of desiccant was then added and the box left for a further 2 hours. The slides were then transferred to slide boxes containing desiccant, one for each exposure time required, the boxes made light-tight, and stored at 4°C for the required exposure time.

The slides were warmed to room temperature for 2 hours, immersed in Kodak D19 developer for 2 minutes at room temperature, followed by incubation in 1% (v/v) acetic acid, 1% (v/v) glycerol for one minute, 30% (w/v) sodium thiosulphate for 2 minutes, washed extensively with distilled water and air dried.

Sections were stained by incubation in 0.02% (v/v) toluidine blue (Raymond Lamb) and mounted.

Section 2.22. DNA Sequencing.

DNA sequences were determined by means of a modified versions of the chain termination method of Sanger et al. (1977). Subcloning of DNA fragments was performed as described in section 2.16.

2.22.1. Preparation of single stranded M13 bacteriophage DNA.

A fresh colony of E.coli JM103 was picked from a minimal agar plate (Table 2.2), inoculated into 0.5ml prewarmed TY broth (Table 2.2) and incubated at 37°C with vigorous aeration for 90 minutes. 200ul of this culture were then added to 2ml prewarmed TY broth containing a fresh M13 plaque. This culture was incubated with vigorous aeration at 37°C for 7 hours.
The culture was placed in two 1.5ml Eppendorf tubes and bacterial debris removed by centrifugation in an Eppendorf microfuge for 15 minutes at 4°C. The supernatant was transferred to a fresh tube and mixed with 200ul 2.5M sodium chloride, 20% (w/v) polyethylene glycol 8000 (Sigma). After incubation at 4°C for 4-16 hours precipitated bacteriophage particles were recovered by centrifugation in an Eppendorf microfuge at 4°C for 10 minutes, the supernatant completely removed, and the pellet air dried for 15 minutes at room temperature. Bacteriophage were resuspended in 100ul TE, samples pooled, extracted with 100ul Tris-saturated phenol, pH7.0, and after centrifugation for 2 minutes at 4°C the

aqueous phase removed to a clean tube. DNA was precipitated by the addition of 0.1 volumes 2M sodium acetate, pH5.5, and 2 volumes of propan-2-ol.

Following incubation at -20°C for one hour and -70°C for 10 minutes, DNA was collected by centrifugation in an Eppendorf microfuge for 10 minutes at 4°C, the pellet washed with 500ul propan-2-ol, air dried, resuspended in 15ul TE, and stored at -20°C.

Integrity of DNA was checked by running 1ul of the preparation on a 0.8% (w/v) agarose gel (section 2.15.1). Yields of approximately 15ug were routinely obtained.

2.22.2. Selective screening of recombinant M13 templates.

The orientation of subcloned fragments to be sequenced was determined by analysis of the products of the T-termination reaction of the dideoxy sequencing method.

Both annealing and sequencing reactions were slightly modified from those given in section 2.22.3. The annealing reaction mix consisted of 1ul sterile water, 0.5ul 100mM Tris-HCl, pH7.5, 70mM magnesium chloride, 2.5ul template DNA, 0.5ul M13 primer (Table 2.7). Annealing was performed as in section 2.22.3.

To 2ul of the annealed mix was added 2.25ul dT^0/ddT mix (Table 2.7) containing 25uCi [α^{35}S] dATP (Amersham International >400Ci/mmol). Initiation of reaction and subsequent manipulations were performed as described in section 2.22.3.

Table 2.7

Reagents for dideoxy sequencing using Klenow fragment of E.coli DNA polymerase.

M13 single stranded primer (17mer) purchased from P.L. Pharmacia (cat.no. 27-1534).

Klenow fragment E.coli DNA polymerase I was purchased from P.L. Pharmacia (cat.no. 27-0928) and diluted in distilled water immediately prior to use.

All nucleoside triphosphates were purchased from P.L. Pharmacia and stored as 10mM solutions in TE at -20°C.

N° mixes

	A°	C°	G°	T°
10mM dCTP	10ul	-	10ul	10ul
10mM dGTP	10ul	10ul	-	10ul
10mM dTTP	10ul	10ul	10ul	-
0.5mM dCTP	-	10ul	-	-
0.5mM dGTP	-	-	10ul	-
0.5mM dTTP	-	-	-	10ul
1x TE	770ul	580ul	580ul	580ul

Stored at -20°C.

Table 2.7 cont.

ddN mixes

	ddA	ddC	ddG	ddT
10mM ddATP	0.67ul	-	-	-
10mM ddCTP	-	2ul	-	-
10mM ddGTP	-	-	6ul	-
10mM ddTTP	-	-	-	10ul
1x TE	199.3ul	198ul	194ul	190ul

Stored at -20°C

N° /ddN mixes

50ul of each of the appropriate N° and ddN solutions were mixed and stored at -20°C.

Chase dNTP mix

5ul 10mM dATP

5ul 10mM dCTP

5ul 10mM dGTP

5ul 10mM dTTP

80ul 1x TE

Stored at -20°C

Formamide dye mix

99.8% (v/v) deionised formamide

0.03% (w/v) xylene cyanol (BDH)

0.03% (w/v) bromophenol blue (BDH)

20mM EDTA, pH8.0

Stored at room temperature.

2.22.3. DNA sequencing reactions using Klenow fragment of E.coli DNA polymerase I.

The annealing reaction mix (5ul template DNA, 1ul 100mM Tris-HCl, pH7.5, 70mM magnesium chloride, 1ul M13 primer (Table 2.7), 3ul sterile water) in a 1.5ml Eppendorf tube was placed in a boiling water bath for 3 minutes prior to incubation at 65^{o}C for 30-60 minutes.

10uCi [α^{35}S] dATP (Amersham International >400Ci/mmol) were added and, after thorough mixing, 2ul aliquots were dispensed into each of four 1.5ml Eppendorf tubes. 2ul of the appropriate N^{0}/ddNTP mix (Table 2.7) was added to each tube, thoroughly mixed, and the tubes placed in an Eppendorf microfuge. 1 unit Klenow fragment of E.coli DNA polymerase I (Table 2.7) was added to the wall of each tube and the reaction initiated by brief centrifugation.

Following incubation at 25^{o}C for 20 minutes 2ul dNTP chase mix (Table 2.7) were added to each tube and the reaction allowed to continue for a further 20 minutes at 25^{o}C. The reaction was terminated by the addition of 4ul formamide dye mix (Table 2.7).

2.22.4. DNA sequencing reactions using modified T7 DNA polymerase.

Determination of DNA sequence using the modified T7 DNA polymerase, Sequenase (United States Biochemical Corporation),

was carried out essentially as described by Tabor and Richardson (1987), with the modifications recommended by the manufacturers.

The annealing mix of 5ul template DNA (section 2.22.1), 2ul annealing mix (Table 2.8) and 1ul M13 primer (Table 2.7), in a 1.5ml Eppendorf tube, was incubated at 65^{o}C for 2 minutes and then placed at 25^{o}C for 30 minutes.

To this was added 1ul 0.1M dithiothreitol, 2ul labelling nucleotide mix (Table 2.8), 0.5ul [α^{35}S] dATP (Amersham International >400 Ci/mmol), mixed well, and the reaction initiated by the addition of 2ul of Sequenase enzyme diluted 8 fold in ice-cold TE immediately prior to use.

Following incubation at 25^{o}C for 10 minutes, 3.5ul aliquots of the reaction mix were added to each of four 1.5ml Eppendorf tubes containing the appropriate dNTP termination mix (Table 2.8), prewarmed to 37^{o}C. After incubation for 5 minutes at 37^{o}C, reactions were terminated by the addition of 4ul formamide dye mix (Table 2.7).

Table 2.8

Reagents for dideoxy sequencing using modified T7 DNA polymerase.

All reagents were purchased from United States Biochemical Corporation.

Annealing buffer.

200mM Tris Hcl, pH7.5

100mM Magnesium chloride

250mM sodium chloride

Stored at -20^{o}C.

Labelling Nucleotide Mix.

7.5uM dGTP

7.5uM dCTP

7.5uM dTTP

Stored at -20^{o}C

Termination nucleotide mixes.

80uM of each dNTP in 50mM sodium chloride with 8uM of the appropriate ddNTP for each of T,C,A and G terminations.

Section 2.23. Transfer of proteins from SDS-PAGE to nitrocellulose membranes.

Proteins were transferred from SDS-PAGE gels (section 2.15.4) to nitrocellulose by use of a semi-dry electroblotter (Sartorius).

Both anode and cathode were thoroughly rinsed with distilled water. Six sheets of 3MM paper (Whatman) soaked in 0.3M Tris, 20% (v/v) methanol, pH10.4, were placed on the anode, followed by three sheets of 3MM soaked in 25mM Tris, 20% (v/v) methanol, pH10.4. A nitrocellulose membrane (Schleicher and Schuell), prewetted with distilled water, was then placed on the filter paper followed by the gel and nine sheets of 3MM soaked in 25mM Tris, 40mM 6-amino-n-hexanoic acid (Sigma), 20% (v/v) methanol, pH9.4. At all stages care was taken to ensure that no air bubbles were trapped. The graphite cathode was placed on top of the "sandwich" and transfer carried out using $0.8mA/cm^2$ for 75 minutes at room temperature.

Section 2.24. Raising polyclonal antisera.

Preparation of the Immunogen.

Proteins transferred to nitrocellulose (section 2.23) were detected by incubation of the filter in Ponceau S solution (Sigma) for 10 minutes with gentle agitation at room temperature. The filter was washed thoroughly with 5% (v/v)

acetic acid and the area containing the galactosidase-fusion protein excised with a scalpel blade. These nitrocellulose fragments were dried briefly under vacuum, dissolved in the minimum volume of dimethylsulphoxide (BDH), and stored at 4°C.

Immunisation of Rabbits.

Up to 250ul (100ug) of fusion protein solution was mixed thoroughly with an equal volume of Freunds complete adjuvant (BRL Gibco) and injected intramuscularly to the thigh. A second thigh injection was performed 7 days later, again using 250ul of fusion protein solution. After a further 5 weeks multiple subcutaneous injections were performed using a total of 250ul fusion protein prepared as before.

Serum was removed for analysis three days after the final injections and subsequently every second day until the antibody titre began to fall.

Blood sampling and preparation of serum.

Blood sampling was achieved by means of the ear bleed procedure, approximately 20-30ml being removed at each bleed. Blood was allowed to clot at room temperature for 90 minutes prior to incubation at 4°C for 16 hours. The clot was pelleted by centrifugation at 6000rpm for 10 minutes at 4°C in a Sorvall RT6000B centrifuge, the serum decanted, and stored in small aliquots at -20°C.

Section 2.25. Western blot analyses.

Proteins were transferred to nitrocellulose as described in section 2.23. The filter was incubated in phosphate buffered saline (PBS) containing 3% (w/v) bovine serum albumin with constant agitation at $4^{o}C$ for 12-16 hours. Following a brief wash with PBS, the filter was incubated in PBS, 3% (w/v) bovine serum albumin, containing an appropriate dilution of test antiserum, for 2 hours with constant agitation at room temperature.

The filter was washed extensively in PBS, 0.1% (v/v) Triton X-100 (BDH) at room temperature prior to incubation for 2 hours with constant agitation at room temperature with PBS, 3% (w/v) bovine serum albumin containing 10^{2} dilution of peroxidase conjugated anti-mouse (DAKO p161) or $5x10^{2}$ dilution of peroxidase conjugated anti-rabbit (ICN biochemicals 61-202) antibody.

Following extensive washing with PBS, 0.1% (v/v) Triton X-100 and a brief wash in PBS alone, the filter was incubated in PBS, 0.05% (v/v) 4-chloro-1-naphthol (Sigma), 0.0002% (v/v) hydrogen peroxide (BDH) at room temperature. Once the colour had developed the filter was washed with distilled water and stored at $4^{o}C$.

Section 2.26. Immunocytochemical staining.

Expression of the stage-specific mouse embryonic antigen SSEA-1 (Solter and Knowles, 1978) in mammalian cell lines was determined by immunocytochemical staining using horseradish peroxidase.

Cell monolayers (section 2.3) were fixed by incubation for one minute at room temperature in a fresh solution of 50% (v/v) methanol, 50% (v/v) acetone. Following a brief wash with phosphate buffered saline (PBS), cells were incubated in PBS containing 10% (v/v) fetal calf serum (Gibco), 10% (w/v) bovine serum albumin (Sigma) and a 10^3 dilution of SSEA-1 monoclonal antibody (MC 480) (kindly provided by Dr. D. Solter, Wistar Institute, Philadelphia) for 2 hours at room temperature.

Cells were washed extensively with PBS prior to incubation in PBS containing 10% (v/v) fetal calf serum, 1% (w/v) bovine serum albumin and a 10^2 dilution of peroxidase conjugated rabbit anti-mouse antiserum (DAKO p161) for a further 2 hours at room temperature.

After further washes with PBS cells positive cells were visualised by incubation in PBS supplemented with 1% (v/v) o-dianisidine (Sigma), 0.003% (v/v) hydrogen peroxide (BDH) for 20 minutes at room temperature.

Chapter 3. Selection of cDNA clones suitable for studies of gene regulation in embryonic stem cells.

Section 3.1. Introduction.

Section 3.1.1. Strategies for the isolation of differentially regulated genes.

As discussed earlier (sections 1.2 and 1.3), murine embryonal carcinoma cells can serve as a suitable in vitro model for the pluripotential stem cells of the early mouse embryo and can be used effectively in the study of transcriptional regulation of gene expression.

Current models of transcriptional regulation in pluripotent embryonic stem cells are largely based on the use of viral transcriptional regulatory elements and there is a need for cellular genes which are clearly down-regulated on chemical induction of EC cell differentiation. To this end, the approach of differential screening of cDNA libraries prepared from EC cells, offers the opportunity not only to isolate clearly regulated genes, but also facilitates the isolation of novel regulated genes.

However, whilst genes must be available which show regulation in the EC cell system, such regulation must also be seen during normal mouse development if regulatory studies are to be relevant. Moreover, analyses of the function of a novel developmentally regulated gene may further understanding of the processes responsible for orderly mammalian development.

Given the interest in understanding the regulation of gene expression in embryonic stem cells, a cDNA library constructed

in the bacteriophage lambda vector gt10 (Huynh et al., 1985) from RNA of the pluripotential EC cell line PCC3 has been screened in two instances with the aim of isolating cDNA clones representing genes the expression of which is repressed upon in vitro differentiation of PCC3 cells.

The first approach followed from the observations of Murphy et al. (1983) that a B2 repetitive element, isolated in a cDNA clone of the murine H-2D^{d} MHC gene (Brickell et al., 1983; Brickell et al., 1985), hybridised to high molecular weight RNA species in embryonic stem cells but not in embryoid bodies produced on in vitro differentiation of that cell line. A similar pattern of expression of B2-containing transcripts was seen on in vitro differentiation of F9 EC cells to parietal endoderm (Murphy et al., 1983). Thus, it appeared that this B2 element could be used as a hybridisation probe for the isolation of cDNA clones representing transcripts the abundance of which decreased on differentiation of embryonal carcinoma cells. On this basis, a cDNA library prepared from PCC3 cells was screened with the B2 repetitive element from the MHC cDNA clone (A.R. Cooper and C.J. Watson, unpublished data).

The second approach made use of radiolabelled cDNA representative of all the mRNA species in a given cell type. Such a differential screen was performed on a cDNA library prepared from PCC3 cells using cDNA probes prepared from PCC3 cells or PCC3 cells treated with retinoic acid for 10 days (Gorman et al., 1985a; Watson, 1986).

In both cases a number of cDNA clones were isolated as bacteriophage, the plaques of which reproducibly hybridised with the B2 probe or reproducibly gave a lower hybridisation signal with a radiolabelled cDNA probe from differentiated PCC3 cells than with a probe prepared from PCC3 cells (Watson, 1986). An analysis of these cDNA clones was initiated in order to determine which, if any, were suitable for studies of transcriptional regulation in early embryonic stem cells.

Section 3.1.2. Criteria for genes suitable for transcriptional analyses.

As discussed earlier (section 1.3), embryonal carcinoma cells are the pluripotential stem cells of teratocarcinomas. By definition, EC cells are highly tumourigenic when transplanted to ectopic sites in vivo whilst the cell types generated on in vitro differentiation have lost this tumourigenic capacity. Thus, clones isolated in a differential screen between undifferentiated and differentiated EC cells will include those representing genes which respond to the loss of the transformed phenotype, in addition to genes the expression of which is sensitive to the loss of the stem cell phenotype. Given that this work is concerned with the regulation of gene expression during embryogenesis, it is essential to eliminate from the analysis any cDNA clones isolated as a consequence of the loss of the transformed phenotype.

Differential screening of cDNA libraries with radiolabelled cDNA probes results in the isolation of clones representing transcripts the steady-state levels of which differ between the two cell types concerned. Such changes in the level of stable cytoplasmic RNA could be due to transcriptional and/or post-transcriptional mechanisms of regulation. Indeed, it has been observed that the down regulation of c-myc and p53 in differentiating F9 cells is exercised at the post-transcriptional level (Dony et al., 1985). Clearly, for any clone to be of use in studies of transcriptional regulation it is essential to establish that the change in steady-state RNA levels between the two cell types can be attributed to alteration in the rate of transcription of that gene. Transcription rates can be measured by incubating isolated nuclei with radiolabelled ribonucleoside triphosphates and annealing the labelled RNA species to an excess of the cloned cDNA immobilised on nitrocellulose (McKnight and Palmiter, 1979; Israel and Whitelock, 1984).

Perhaps the most important criterion in determining the usefulness of any gene for transcriptional studies in an in vitro cell system is that the regulated expression observed in vitro is a suitable representation of the expression pattern of that gene in vivo. Analysis of the expression pattern of the gene during normal mouse development, in addition to validating the regulation seen in the EC cell system, may reveal additional temporal and/or cell type specific patterns of expression, so expanding the variety of cell types in which

the regulatory mechanisms operating on the gene can be analysed.

The ability of a cDNA clone to meet the criteria outlined so far can be established by means of Northern blot analyses, using radiolabelled cDNA insert as a hybridisation probe to determine differential expression in a variety of cell and tissue types, along with transcriptional run-on analyses in isolated nuclei.

An essential feature of any Northern blot analysis designed to establish differential expression patterns is the use of an appropriate control probe to check both quantitative and qualitative features of the experiment. Whilst it is difficult to identify the ideal control gene, i.e. one which is expressed at equivalent levels in all cell types under all growth conditions, nuclear genes which encode proteins essential for basic cellular functions were considered useful. Many workers use actin as a control probe. In this work, control probes homologous to murine hypoxanthine phosphoribosyl transferase (HPRT) or human αtubulin (Table 2.3) were used to ensure equivalent amounts of intact RNA were present in all lanes in Northern blot analyses. In some cases, it may be desirable to use probes from genes the expression of which is characteristic of a given cell type, such as laminin to confirm differentiation of EC cells to parietal endoderm (Barlow et al., 1984).

It was also considered important that the cDNA clones isolated should represent transcripts of low-copy number genes. Some

classes of endogenous retroviral elements are known to exhibit differential regulation in the EC cell system, including Etn (Brulet et al., 1983) and intracisternal A particles (Hojman-Montes de Oca et al., 1983; Howe and Overton, 1986). Whilst viral enhancer/promoter elements have been used to develop models for the mechanisms of regulation of gene expression in embryonic stem cells (section 1.3), it is desirable to test these models on the regulatory elements of a low-copy number murine cellular gene.

Moreover, nucleotide sequences of different members of the same family of endogenous retroviral elements often show considerable divergence (Keshet and Itin, 1982; Carter et al., 1983; Burt et al., 1984) and not all members of a given family of elements may be transcriptionally active. In addition, the LTRs of at least some transcriptionally active members of the Etn family of endogenous retroviral-like elements contain relatively weak transcriptional promoter elements (Brulet, pers. comm.). Thus, considerable difficulty may be encountered in attempts to isolate transcriptional regulatory elements of repetitive elements or multigene families.

Also, the absence of any polypeptide encoding capacity in the ETn genome (Sonigo et al., 1987) raises the possibility that such endogenous retrovirus-like elements act as "selfish" or "parasitic" DNA (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Thus, the role, if any, of these elements in normal mouse development is unclear. For these reasons, analysis of cDNA clones corresponding to repetitive elements

was considered undesirable.

Obviously, it was important to determine whether the genes represented by the cDNA clones were novel or had been previously characterised. This was achieved by determining the nucleotide sequence of the cDNA clone. Computer assisted comparisons of the determined sequence with data bases of known DNA and amino acid sequences were then conducted in order to determine if the cDNA clone represented a previously characterised gene.

To determine the cDNA clones representing genes which met the above criteria, a number of clones were used in Northern blot analyses, initially to confirm that the putative regulation seen at the level of bacteriophage plaque hybridisation was genuine. Northern blot analyses were also used to determine the expression pattern of the genes in normal and SV40-transformed fibroblasts, and other EC cell lines, to ascertain which clones represented genes sensitive to the transformed phenotype. Clones which met the criteria outlined were then used in further Northern blot experiments to establish the expression pattern of that gene during normal mouse development.

Section 3.2. Results.

Section 3.2.1. Expression of 1.6U.

1.6U was isolated in a screen of a PCC3 cDNA library as a clone homologous with the B2 repetitive element located in the 3' untranslated region of a H-2D^d class I MHC gene (Brickell et al., 1983; Brickell et al., 1985; A.R. Cooper and C.J. Watson, unpublished data). The 1.6kb EcoRI insert fragment of this clone was subcloned into pBR325 to create the plasmid p1.6. Preliminary restriction mapping and Southern blot analyses located the B2 homologous region to a 200bp PstI-EcoRI fragment at one end of the insert (fig. 3.1). The 1.1kb PstI fragment of this clone was radiolabelled by the method of random priming (section 2.17.1) and used for Northern blot analyses.

The probe was found to hybridise to two abundant transcripts in PCC3 cells of approximately 2.1kb and 3.0kb. These transcripts were also detected in similar abundance in PCC3 cultures induced to differentiate either by retinoic acid treatment or suspension culture to produce embryoid bodies (fig. 3.2a). Northern blot analyses of the EC cell lines PCC4 and F9, along with the differentiated derivatives of such cells (section 2.3), showed the two 1.6U homologous transcripts to be present at similar levels in all 4 cell types (fig. 3.2a), whilst the small B2 RNA species were down regulated on differentiation in agreement with the

Figure 3.1. Restriction Map of p1.6.

The 1.6kb cDNA clone 1.6U in the bacteriophage lambda vector gt10 was subcloned as a single EcoRI fragment into the plasmid vector pBR325 to give the plasmid p1.6 (A). The nucleotide sequence of two regions of this clone, MR and HP, were determined (section 3.2.3). The 1.1kb PstI fragment was subcloned into the plasmid vector pBSM13$^+$ to give the plasmid pBS1.6U (B).

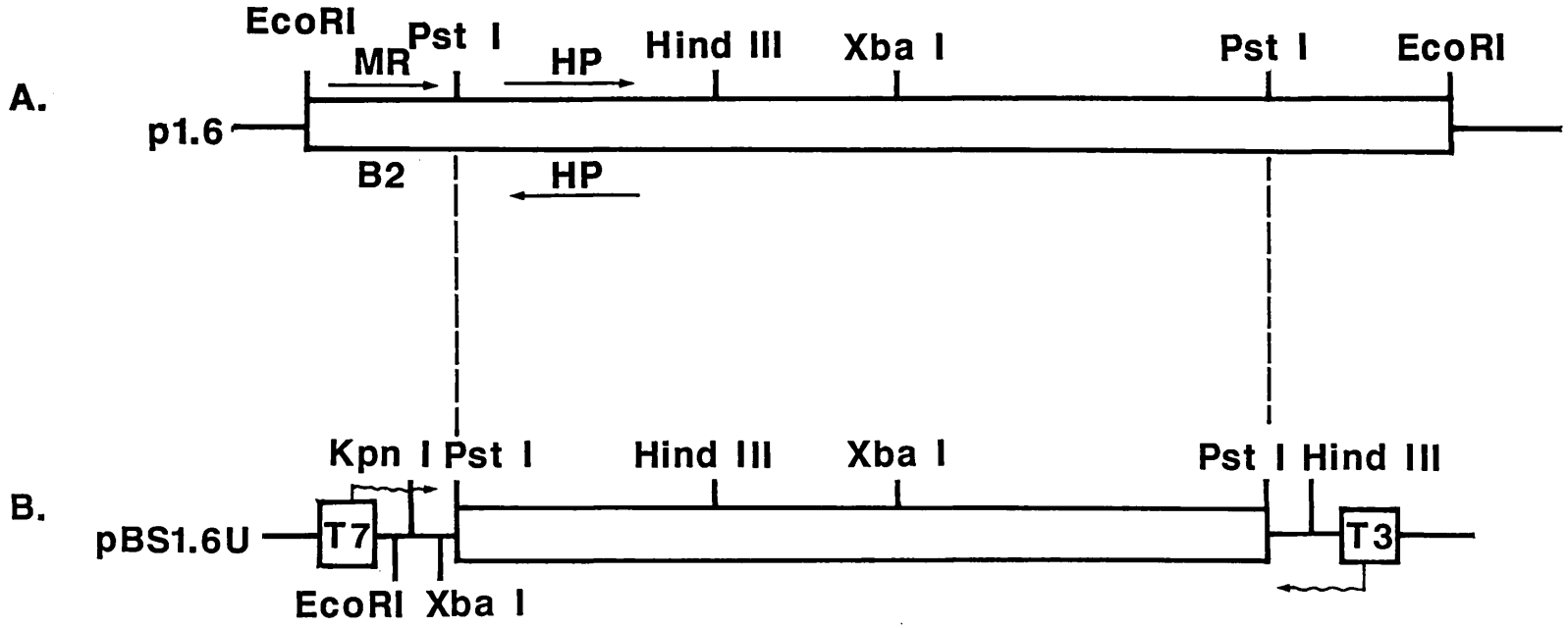
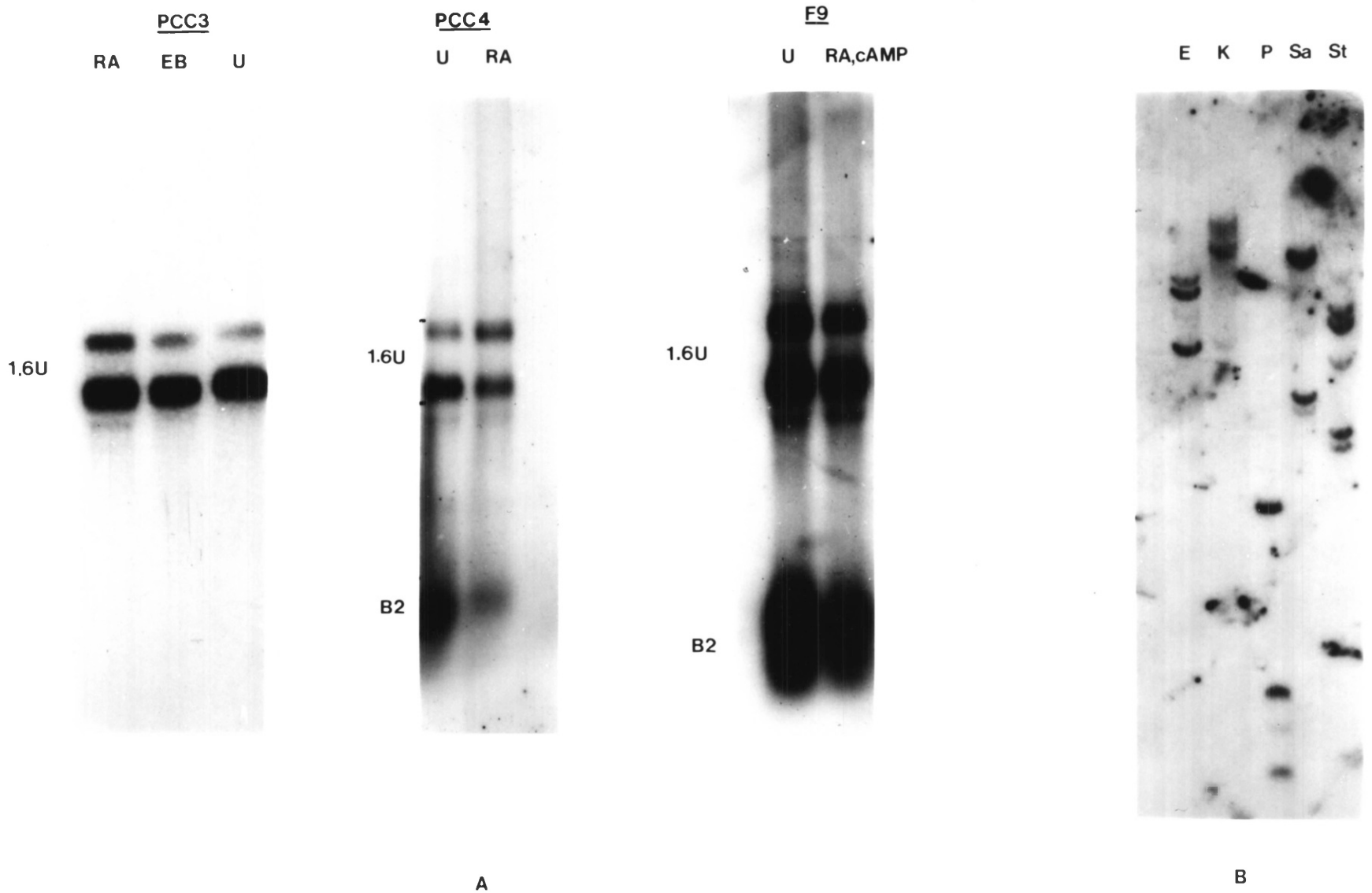

A.
EcoRI
MR
Pst I
HP
Hind III
Xba I
Pst I
EcoRI
p1.6
B2
HP
B.
Kpn I
Pst I
Hind III
Xba I
Pst I
Hind III
pBS1.6U
T7
T3
EcoRI
Xba I

Figure 3.2. A. Northern blot analysis of 1.6U expression in EC cells.

2ug of polyadenylated cytoplasmic RNA from PCC3 cells (U), PCC3 embryoid bodies (EB), retinoic acid-treated PCC3 cells (RA) (provided by C.J. Watson), PCC4 cells (U), PCC4 cells treated with retinoic acid for 7 days (RA), 3ug polyadenylated cytoplasmic RNA from F9 cells (U), or F9 cells treated with retinoic acid, cyclic AMP and IBMX for 5 days (RA,cAMP), were probed with oligolabelled EcoRI fragment of p1.6 (specific activity $9x10^7$cpm/ug; $9x10^6$ cpm used per filter). Filters were washed to 0.2xSSC, 0.1% (w/v) SDS at 55^oC and exposed for 2 hours (PCC3), or 16 hours (PCC4 and F9). PCC4 and F9 filters were subsequently probed with oligolabelled EcoRI-HinDIII fragment of p64B2 (Table 2.3) (specific activity $2x10^7$cpm/ug; 10^6cpm/ml used) and washed to 0.2xSSC, 0.1% (w/v) SDS at 55^oC and exposed for 16 hours.

B. Southern blot analysis of 1.6U.

10ug of EcoRI (E), KpnI (K), PstI (P), SacI (Sa) and StuI (St) digested PCC3 genomic DNA restriction fragments were probed with radiolabelled RNA synthesised in vitro by T3 polymerase using a template of pBS1.6U linearised with KpnI (specific activity $1.6x10^9$ cpm/ug; $1x10^6$cpm/ml used). The filter was washed to 0.1xSSC, 0.1% (w/v) SDS at 65^oC and exposed for 7 days.



A

B

observations of Murphy et al. (1983). Further analysis of the expression pattern of 1.6U showed high levels of expression of both transcripts in Balb/c adult liver and spleen as well as the murine SV40-transformed cell line SV3T3 C138 (data not shown). It is unclear whether the two 1.6U homologous transcripts represent products of two distinct loci or differentially initiated/terminated or spliced transcripts from a single transcription unit. The fact that both transcripts are abundant in polyadenylated cytoplasmic RNA suggests that they do not represent spliced and unspliced precursor RNA species.

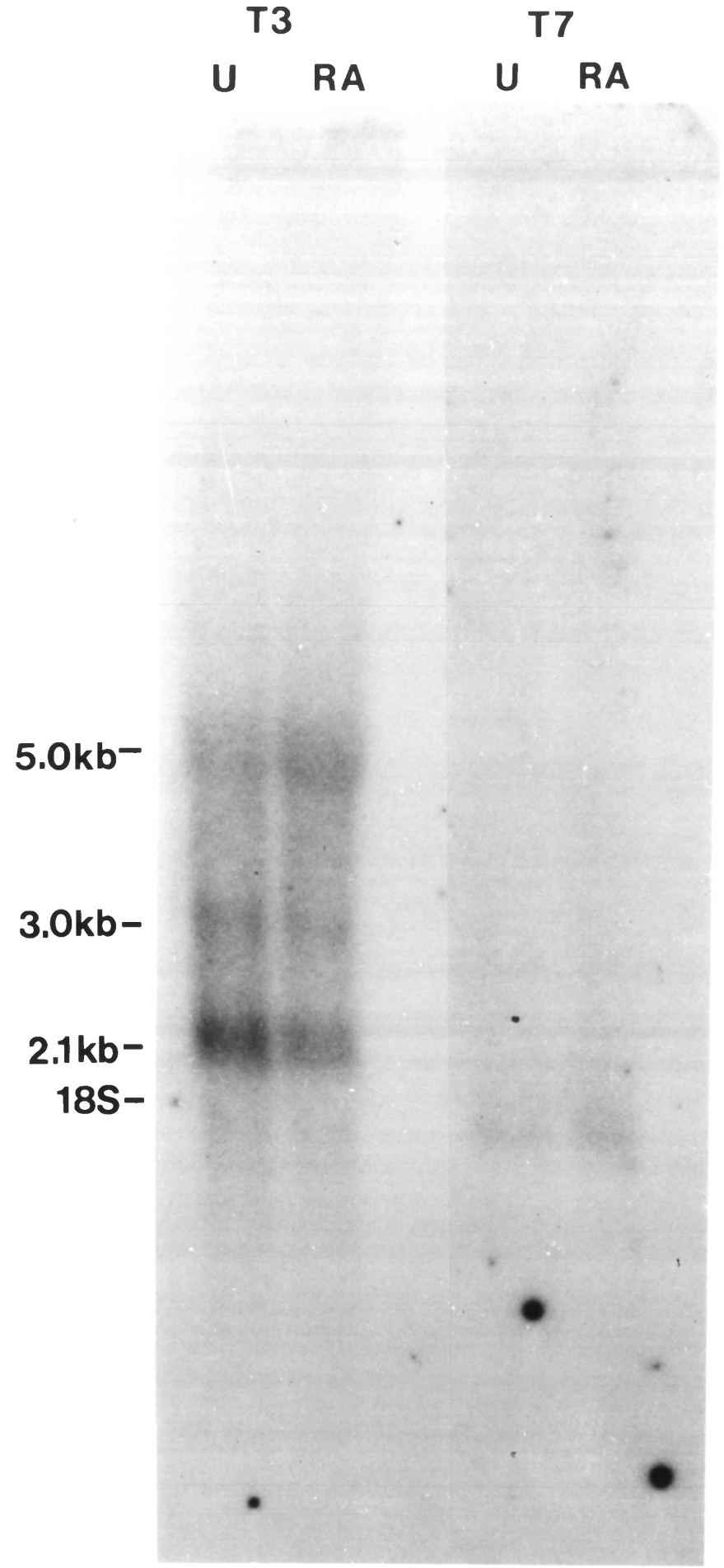
The 1.1kb PstI fragment of p1.6 was subcloned into the plasmid vector pBSM13$^+$ to produce the plasmid pBS1.6U (fig. 3.1b). A strand-specific RNA probe was synthesised from this plasmid for use in Southern blot analysis of mouse genomic DNA. This showed the 1.6U gene(s) to be present at a low copy number (1-3 genes) in the mouse genome (fig. 3.2b).

Section 3.2.2. Transcriptional orientation of the p1.6 cDNA clone.

In order to determine the orientation of the B2 repeat in the p1.6 cDNA clone relative to the direction of transcription, strand-specific RNA probes were generated using T3 or T7 polymerases on linearised templates of pBS1.6U (fig. 3.1b) and used in Northern blot analyses. As shown in fig. 3.3, only one

Figure 3.3. Transcriptional orientation of p1.6.

10ug of total cellular RNA from PCC3 cells (U) or PCC3 cells grown in retinoic acid containing medium for 10 days (RA) were probed with radiolabelled strand-specific RNA synthesised *in vitro* using T3 polymerase on a template of pBS1.6U linearised with *Eco*RI (T3) or T7 polymerase on a template of pBS1.6U linearised with *Hin*DIII (T7) (specific activities 2.4×10^9cpm/ug and 7.2×10^8cpm/ug respectively; 10^6cpm/ml used in each case). In both cases the filter was washed to 0.1xSSC, 0.1% (w/v) SDS at 65°C and exposed for 3 days.


T3
T7
U
RA
U
RA
5.0kb
3.0kb
2.1kb
18S

strand hybridised to the characteristic 2.1kb and 3.0kb mRNAs in PCC3 cells. The other strand hybridised weakly to a transcript migrating marginally faster than the abundant 2.1kb 1.6U mRNA. This hybridisation may be due to weak homology of the sense strand probe with 18S ribosomal RNA but clearly there is no hybridisation to the major 2.1kb and 3.0kb 1.6U transcripts. Given this transcriptional sense of the cDNA clone, it is apparent that the B2 element lies at the 5′ end. An additional transcript, estimated to be 5kb in size, was detected with the anti-sense RNA probe. Given that total cellular RNA was used in this experiment, and that the 5Kb transcript was not detected in polyadenylated, cytoplasmic RNA (fig. 3.2), it is likely that it represents an unspliced precursor of the 1.6U transcription unit(s).

Section 3.2.3. DNA sequence analysis of p1.6.

The nucleotide sequences of two regions of this cDNA clone, MR and HP (fig. 3.1a), were determined by the chain termination method using Klenow fragment of E.coli DNA polymerase I (section 2.22.3) and are presented in figure 3.4.

The 1.6-MR sequence was closely homologous to the lower strand of the consensus sequence of B2 repetitive elements (Kramerov et al., 1985; fig.3.4b). This confirmed the presence of a B2 repetitive element in the cDNA clone p1.6. Moreover, given the known transcriptional orientation of the 1.6U gene (section

Figure 3.4. Partial nucleotide sequence of p1.6.

A. The nucleotide sequence of the sense strand of the HP region of p1.6 (fig. 3.1a) is shown, along with the predicted amino acid sequence of the single open reading frame of this sequence. This sequence is a consensus derived from that of both the 560bp EcoRI-HinDIII and 1.1kb PstI-PstI restriction fragments of p1.6U cloned into M13mp8.

B. The nucleotide sequence of the sense strand of the 1.6-MR region of p1.6 (fig. 3.1a) is compared with the lower strand of the B2 consensus sequence, B2 inv (Kramerov et al., 1985). The 1.6-MR sequence file was derived from the 560bp EcoRI-HinDIII restriction fragment of p1.6 (fig. 3.1a) subcloned in M13mp9. * = mis-matched nucleotide.

A

```
                  15                  30                  45
C AGT CTT TTC TCG AAG TTC TTT GAT TTT AGC ACC TTT AAC ACC AAT TAT TCC
  Ser Leu Phe Ser Lys Phe Phe Asp Phe Ser Thr Phe Asn Thr Asn Tyr Ser

         60                  75                  90
  TCC TGC CAG ACT CTG ATG AAT CAA CAG TCT CAA CTC GCA ATC AAG TCA TTC
  Ser Cys Gln Thr Leu MET Asn Gln Gln Ser Gln Leu Ala Ile Lys Ser Phe

105                 120                 135
  CTT TAT AAT GTT GGT AAT TTA AGC ATC CAC AGC AGC AGA GTC GAG CG
  Leu Tyr Asn Val Gly Asn Leu Ser Ile His Ser Ser Arg Val Glu
```

B

```
                                  *                   *       ****
1.6-MR    2 ATTTATTATATGTAAGTACATTGTAGCTGTCTTCAGACATTCCAGAGNNGNGGG
B2 inv   11 ATTTATTATATGTAAGTACACTGTAGCTGTCTTCAGACACTCCAGAG   AGGG

         *                                                     *
1.6-MR   CGTCAGATCTCGTTACGGATGGTTGTGAGCCACCATGTGGTTGCTGGGATTTGA CT 109
B2 inv   AGTCAGATCTCGTTACGGATGGTTGTGAGCCACCATGTGGTTGCTGGGATTTGAACT 119
```

3.2.2), it was apparent that the lower strand of this B2 element would be present in 1.6U mRNA.

The sense strand of the 1.6-HP DNA sequence was found to contain a single open reading frame (fig. 3.4a). Comparison of both nucleotide and putative amino acid sequences with those in the October 1987 and July 1987 releases of EMBL data bases (section 6.1.2) failed to reveal any homologous sequences (data not shown). Thus, it was concluded that 1.6U represents a gene with a capacity to encode a novel polypeptide.

Section 3.2.4. Expression of λ22 in murine embryonal carcinoma cells.

The cDNA clone λ22 was also isolated on differential screening of the PCC3 cDNA library (Gorman et al., 1985; Watson, 1986). The EcoRI restriction fragment containing the cDNA insert of this clone was subcloned into pSP65 to give the recombinant plasmid pSP2265 (fig. 3.5).

The oligolabelled EcoRI cDNA restriction fragment of p2265 was used as a probe for Northern blot experiments with RNA isolated from three murine embryonal carcinoma cell lines, PCC3, PCC4, and F9, along with the differentiated derivatives of each. In order to determine whether the down regulation observed was due to the loss of the transformed phenotype, RNA samples from a normal mouse fibroblast cell line (Balb/c 3T3A31) and from an SV40-transformed derivative (SV3T3 C138)

Figure 3.5. Restriction map of λ22.

The 2.2kb cDNA clone λ22 in the bacteriophage lambda vector gt10 was subcloned as a single EcoRI fragment into the plasmid vector pSP65 to generate the plasmid p2265. The nucleotide sequence of the region 9-22 of this cDNA clone was determined as described in section 3.2.5.

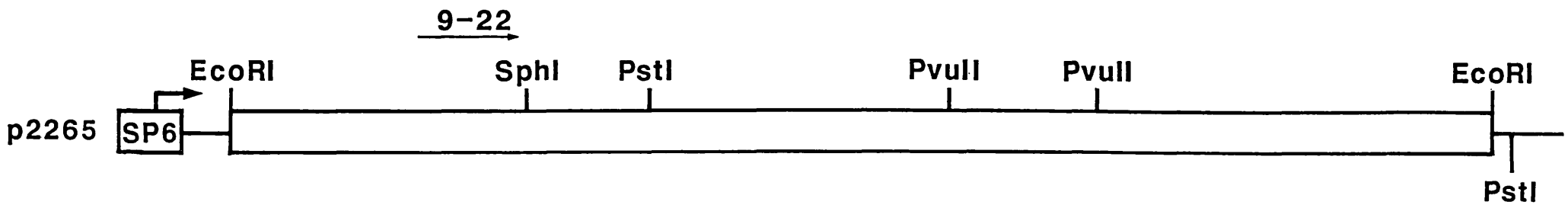

9-22
EcoRI
SphI
PstI
PvuII
PvuII
EcoRI
p2265
SP6
PstI

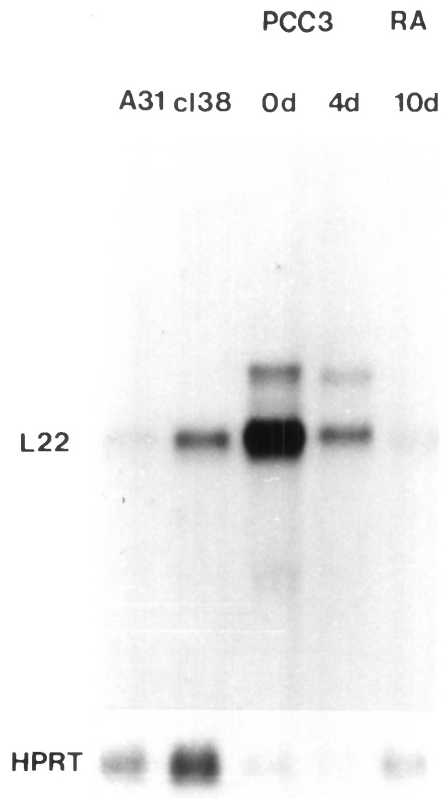
Figure 3.6. Northern blot analysis of λ22 expression in EC cells.

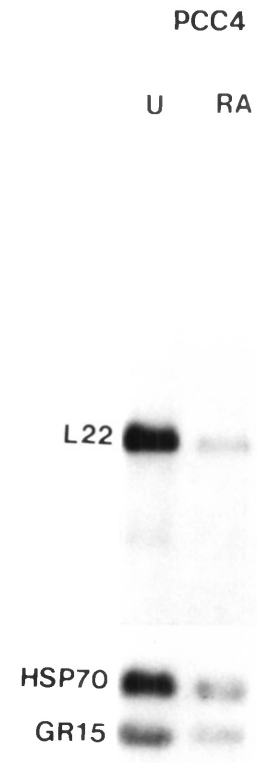
A. 5ug of total cytoplasmic RNA from Balb/c 3T3A31, SV3T3 C138 and PCC3 cells treated with retinoic acid for 0 days, 4 days, or 10 days were probed with oligolabelled EcoRI fragment of λ22 (specific activity $7.6x10^7$cpm/ug; 10^6cpm/ml used) (L22) and oligolabelled PstI fragment of the mouse HPRT cDNA clone, pHPT4 ($7x10^5$cpm/ml used).

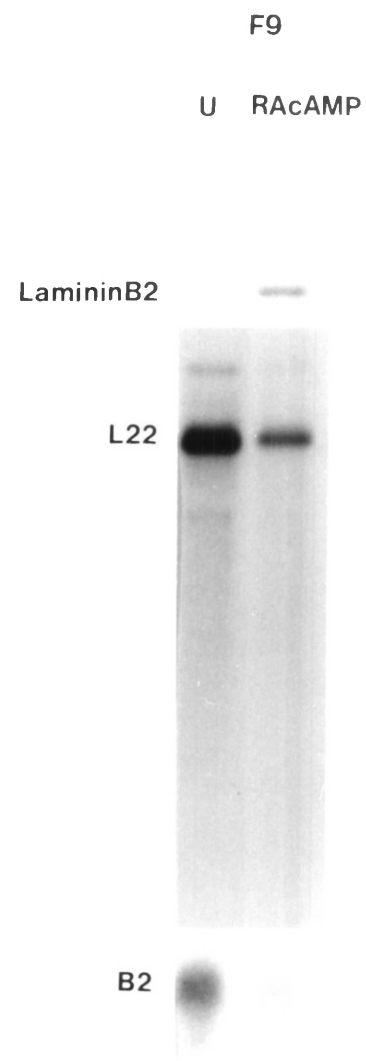
B. The PCC4 filter (fig. 3.2) was probed with oligolabelled EcoRI fragment of λ22 (specific activity $7.6x10^7$cpm/ug; 10^6cpm/ml used) (L22), oligolabelled EcoRI fragment of λGR15 (provided by C.J. Watson) and oligolabelled fragment of HSP70 (provided by A.R. Cooper. Specific activity $2.2x10^7$cpm/ug; $8x10^5$cpm/ml used).

C. The F9 filter (fig. 3.2) was probed with oligolabelled EcoRI fragment of and λ22 (specific activity $7.6x10^7$cpm/ug; 10^6cpm/ml used) (L22), oligolabelled EcoRI-PstI fragment of the laminin B2 chain cDNA clone pPE9 (specific activity $7x10^8$cpm/ug; $3.5x10^6$cpm/ml used). The B2 probe was as described in fig. 3.2. All filters were washed to 0.2xSSC, 0.1% (w/v) SDS at 55°C and exposed for 5 hours (L22 and pPE9) or 16 hours (pHPT4 and GR15).

The hybridisation signal seen above the L22 RNA species in PCC3 and F9 filters is a residual signal from previous hybridisation of these filters with the cDNA clone pGR165 (section 3.2.7).


PCC3
RA
A31 cl38
0d
4d
10d
L22
HPRT


PCC4
U
RA
L22
HSP70
GR15


F9
U
RAcAMP
LamininB2
L22
B2

were included in the analysis.

The λ22 transcript of approximately 5kb, estimated from the rate of migration of GR1 (section 3.2.7) and HPRT mRNA species on the same filter, was down regulated on in vitro differentiation of all three EC cell lines (fig. 3.6). However, the λ22 transcript was also detected at low levels in the mouse fibroblast cell line Balb/c 3T3A31 and was apparently more abundant in the SV40-transformed derivative cell line SV3T3 C138. (The higher molecular weight transcript seen above that of λ22 in fig. 3.6 is residual GR1 signal (section 3.2.7) left as a result of inefficient stripping of the filter prior to rehybridisation).

Should the steady-state level of HPRT mRNA be insensitive to the transformed phenotype, there may be greater amounts of RNA in the SV3T3 C138 track relative to the Balb/c 3T3A31 RNA sample in figure 3.6. Thus, the apparent increase in λ22 RNA levels on SV40 transformation may not be as great as suggested in fig. 3.6.

Nevertheless, this result raises the possibility that expression of λ22 is sensitive to the transformed phenotype, making it possible that some of the regulation observed on in vitro differentiation of EC cells may be attributable to the loss of the transformed phenotype. As such, this would complicate any studies concerning the developmental regulation of this gene in EC cells.

Section 3.2.5. DNA sequence analysis of p2265.

In order to determine whether λ22 represents the transcript of a previously characterised gene, a limited restriction map of the 2.2kb cDNA insert in pSP2265 was determined (fig. 3.5). The 540bp PstI-EcoRI restriction fragment was then subcloned into M13mp9 (section 2.16) and sequenced by the chain termination method using Klenow fragment of E.coli DNA polymerase I (section 2.22.3). The nucleotide sequence determined is shown in fig. 3.7.

Analysis of this sequence indicated that neither strand of this region of p2265 contained a long open reading frame (data not shown). Given that the cDNA clones were generated by the use of an oligo dT primer for first strand cDNA synthesis (Watson and Jackson, 1985), it is likely that this region of the p2265 cDNA clone represents the 3′ untranslated region of the λ22 transcript.

Comparison of this sequence with all nucleotide sequences in the October 1987 release of the EMBL data base (section 6.1.2) failed to reveal any significant homology with previously characterised DNA molecules (data not shown). On this basis, it was concluded that λ22 is a novel gene.

Figure 3.7. Partial nucleotide sequence of p2265.

The nucleotide sequence of the upper strand of the 9-22 region of p2265 (fig. 3.5) is shown.

```
                   20                  40                  60
AAAGTGCAATAGTCACTGCCTTACAGCCAAGTATTAAATGTTTGTCAACTAGTATTAATGCAAAAGTTTAGG

      80                  100                 120                 140
TATTGTAAAATTTTAAAACTTTAACTTCTTAAAAGACAAAAAAGAGAGAAATGTATCCCCGTGCATGCTATT

              160                 180                 200
TAGTGCATTCCCATGCACACTATTAAAGTCTTACCTTTATTTTCAAAATCTAATATTTTAATGTCAAAGAGT

 220                 240                 260                 280
TTAATAAATGCTTTATAGTATCTTAAAGGGATCTACTTATTGGCTTATAGATTAATCCTAAAACAGCTACCT


TATTA
```

Section 3.2.6. Expression pattern of GR4 in murine embryonal carcinoma cells.

GR4 was another clone isolated in the differential screen of the PCC3 cDNA library (Gorman et al., 1985; Watson, 1986). It was shown to exhibit a 2 fold reduction in expression on differentiation of PCC3 cells (Watson, 1986). The same filters used for 22 (section 3.2.4) and GR1 (section 3.2.7) expression analyses were probed with radiolabelled EcoRI cDNA insert fragment of GR4. As shown in fig. 3.8a, the cDNA clone hybridised to a single RNA species of approximately 2kb which was expressed at approximately equivalent levels in both undifferentiated and retinoic acid treated PCC3 cells and normal and SV40-transformed fibroblasts. GR4 showed a small down regulation on F9 differentiation to parietal endoderm (fig. 3.8a) and showed no apparent regulation on PCC4 differentiation in vitro. The significance, if any, of the apparent lower levels of expression in PCC4 cells relative to PCC3 and F9 cell types is unclear since the Northern blot experiments were designed to detect quantitative differences on differentiation of a given EC cell line and not between different EC cell lines.

The radiolabelled EcoRI cDNA fragment of GR4 was used in Southern blot analysis of mouse genomic DNA. The signal intensity of hybridising restriction fragments seen with the GR4 probe was similar to that when a radiolabelled restriction fragment of clone GR1 (section 3.2.7) was used to probe

Figure 3.8. A. Northern blot analysis of GR4 expression in EC cells.

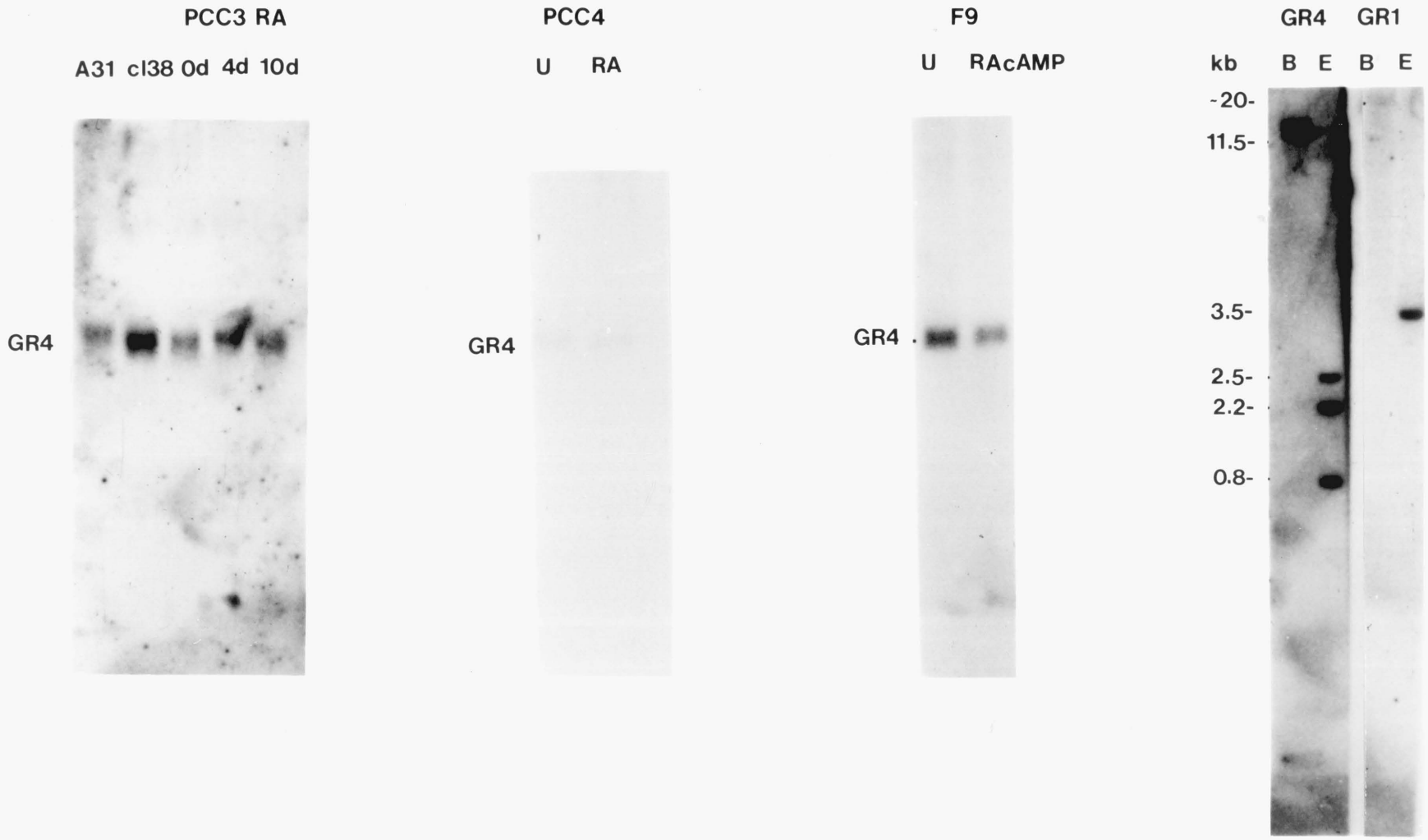
PCC3 (fig. 3.6), PCC4 and F9 (fig. 3.2) filters were probed with oligolabelled EcoRI fragment of GR4 (specific activity $5.6x10^7$cpm/ug; $1.7x10^6$cpm/ml, $1.4x10^6$cpm/ml and $6x10^5$cpm/ml used for PCC3, PCC4, and F9 filters respectively), washed to 0.2xSSC, 0.1% (w/v) SDS at 55°C and exposed for 5 hours (PCC3 filter) or 18 hours (PCC4 and F9 filters).

Other probes, as before (fig. 3.2, fig. 3.6).

B. Southern blot analysis of GR4.

10ug samples of mouse genomic DNA (isolated from the cell line SV3T3 C138) digested with restriction enzymes BamHI (B) and EcoRI (E) were probed with oligolabelled EcoRI fragment of GR4 (specific activity 10^8cpm/ug; $2x10^6$cpm/ml used) (GR4) or oligolabelled EcoRI fragment of pGR165 (section 3.2.7) (specific activity $7.2x10^7$cpm/ug; $2x10^6$cpm/ml used) (GR1). In both cases, filters were washed to 0.1xSSC, 0.1% (w/v) SDS at 65°C, and exposed for 4 days.

The sizes of hybridising fragments were estimated by comparison with the rate of migration of HindIII and EcoRI-HindIII restriction fragments of bacteriophage lambda DNA. kb= kilobases.


PCC3 RA
A31 cl38 0d 4d 10d
GR4
PCC4
U RA
GR4
F9
U RAcAMP
GR4
GR4 GR1
kb B E B E
-20-
11.5-
3.5-
2.5-
2.2-
0.8-
A
B

aliquots of the same genomic DNA restriction fragments (fig. 3.8b). Given that GR1 is known to be present at a low copy number in the mouse genome (section 4.2), it was concluded that GR4 was also a low copy gene.

Section 3.2.7. Expression of GR1 in murine embryonal carcinoma cells.

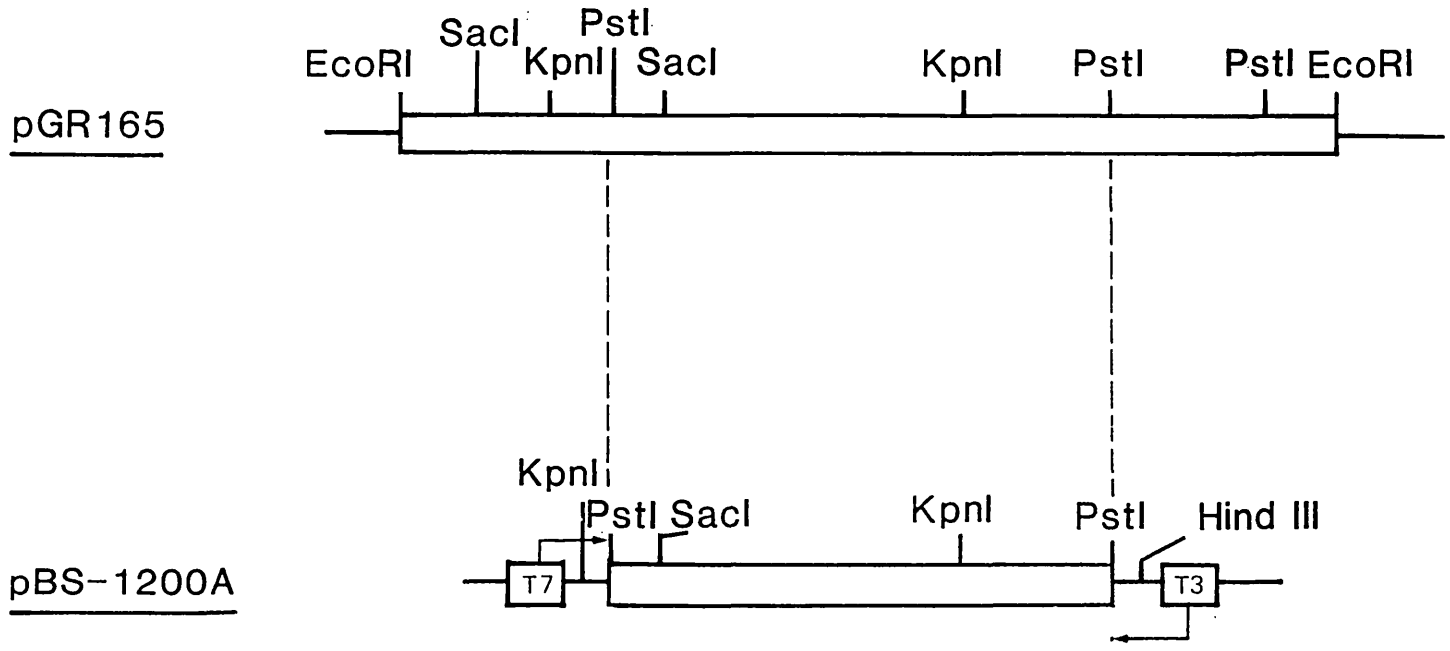
The cDNA clone GR1 was isolated in a differential screen of the PCC3 cDNA library as one of a number of clones representing transcripts the steady-state levels of which decrease on *in vitro* differentiation of this cell line (Gorman *et al.*, 1985a; Watson, 1986). The insert of this cDNA clone was shown to hybridise to an abundant transcript in PCC3 cells which decreased approximately 10 fold after retinoic acid induced differentiation (Watson, 1986).

The 2.5kb cDNA insert was subcloned as a single *Eco*RI fragment into the vector pSP65 to give the plasmid pGR165. The 1.2kb *Pst*I fragment was subcloned into the plasmid vector pBSM13$^+$ to generate the plasmid pBS1200A (fig. 3.9). The entire *Eco*RI fragment was used to probe the EC cell filters used for both 22 (section 3.2.4) and GR4 (section 3.2.6) expression analyses.

As shown in fig. 3.10, the GR1 transcript, estimated to be 6.8kb in size by comparison with the rate of migration of standard marker RNA species (section 2.15.2; data not shown),

Figure 3.9. Restriction map of pGR165.

The 2.5kb cDNA insert of clone GR1 in the bacteriophage lambda vector gt10 was subcloned as a single EcoRI fragment into the vector pSP65 to give the plasmid pGR165. The 1.2kb PstI fragment was further subcloned into the vector pBSM13^{+} to generate the plasmid pBS-1200A.


pGR165
EcoRI
SacI
KpnI
PstI
SacI
KpnI
PstI
PstI
EcoRI
pBS-1200A
KpnI
PstI
SacI
KpnI
PstI
Hind III
T7
T3

was readily detected in PCC3 cells treated with retinoic acid for 0 days or 4 days but was barely detectable in PCC3 cultures treated with retinoic acid for 10 days. This confirmed the data obtained from the original bacteriophage plaque hybridisations of the PCC3 cDNA library, which involved the use of radiolabelled cDNA probes from undifferentiated or 10 day retinoic acid-treated PCC3 cells (Gorman et al., 1985a; Watson, 1986). Moreover, the GR1 transcript was not seen in either normal or SV40-transformed fibroblasts suggesting that the expression of this gene in EC cells is not a general feature of the transformed phenotype. The oligolabelled PstI fragment of pHPT4 (Konecki et al., 1982) was used to indicate the presence of intact HPRT mRNA in all lanes.

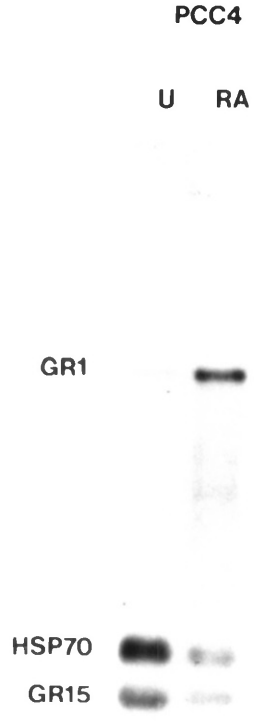
The 6.8kb GR1 transcript was also seen in the EC cell lines PCC4 and F9 (fig. 3.10). However, in contrast to the regulation seen on differentiation of PCC3 cells, an increase in the steady-state level of GR1 mRNA was observed on in vitro differentiation of PCC4 cells whilst differentiation of F9 cells in vitro to parietal endoderm (Strickland et al., 1980) resulted in no change in GR1 expression. Oligolabelled probes homologous to heat-shock protein 70 (HSP70) and GR15 (Watson, 1986) indicated that the increase in the steady-state level of GR1 mRNA on differentiation of PCC4 cells was not an artefact of the Northern blot procedure. Also, the use of a laminin B2 chain cDNA probe (Barlow et al., 1984), along with the observed decrease in sequences homologous to a B2 repetitive element (Murphy et al., 1983), on the F9 filter confirm the

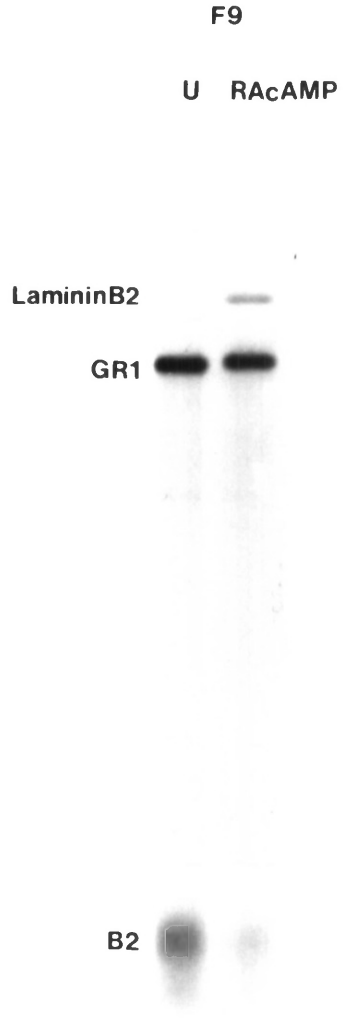
Figure 3.10. Northern blot analysis of GR1 expression in EC cells.

PCC3 (fig. 3.6), PCC4 and F9 (fig. 3.2) filters were probed with oligolabelled EcoRI fragment of pGR165 (specific activities $2x10^8$cpm/ug; $5x10^5$cpm/ml used, $7.7x10^7$cpm/ug; 10^6cpm/ml used and $2.8x10^5$cpm/ml used for each of PCC3, PCC4 and F9 filters, respectively).

All filters were washed to 0.2xSSC, 0.1% (w/v) SDS at 55^oC and exposed for 1 hour (PCC3 filter), 5 hours (PCC4 filter) and 16 hours (F9 filter). Other probes as before (fig. 3.2 and fig. 3.6).


PCC3
RA
A31
cl38
0d
4d
10d
GR1
HPRT


PCC4
U
RA
GR1
HSP70
GR15


F9
U
RA
cAMP
LaminB2
GR1
B2

differentiation of this cell line to parietal endoderm.
Given the observed patterns of regulation of the GR1 transcript in different EC cell lines, and the non-correlation between GR1 expression and the transformed phenotype, it was concluded that this cDNA clone was worthy of further study.

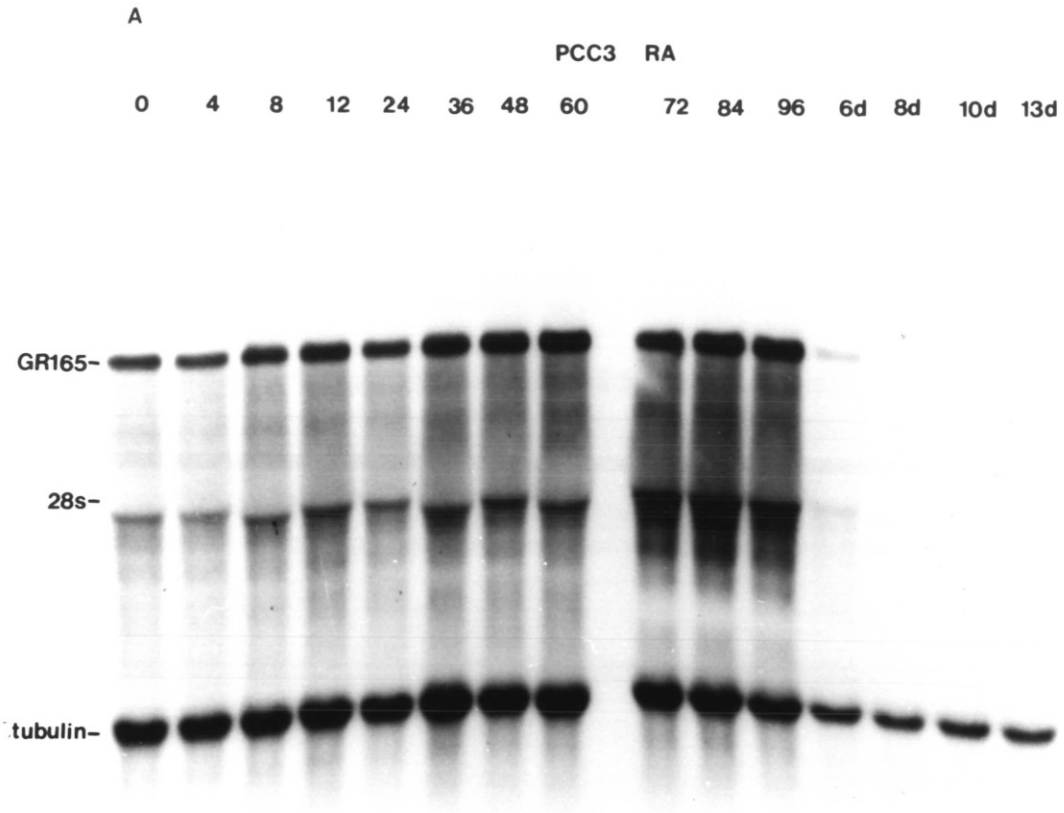
Section 3.2.8. Kinetics of GR1 regulation on *in vitro* differentiation of PCC3 cells.

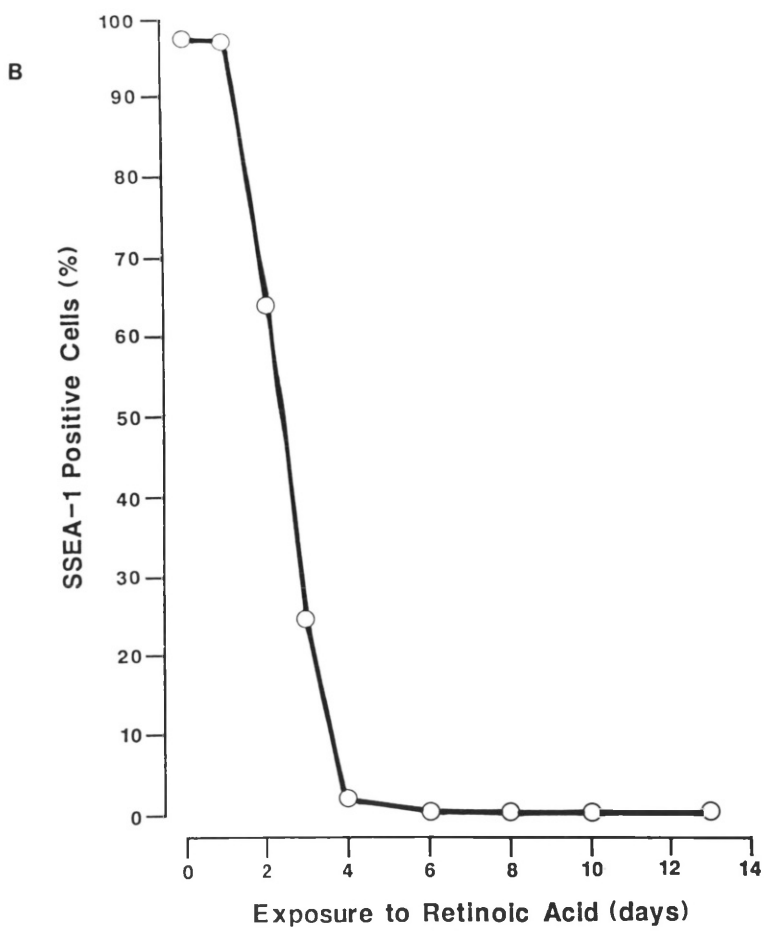
Exposure of EC cells to retinoic acid results in a decreased proliferation rate, loss of the transformed phenotype and altered cell morphology, ultimately resulting in the formation of a terminally differentiated cell type. Although the mechanism by which retinoic acid induces such changes is not fully understood (section 1.2.2), genes the expression of which responds rapidly to retinoic acid treatment would be thought to be involved in earlier steps in the differentiation process than genes the expression of which is altered subsequent to changes in EC cell proliferation and morphology. It was considered of interest to determine the kinetics of GR1 regulation in differentiating PCC3 cells relative to morphological and proliferation rate changes in such cultures, in order to establish at what stage in the differentiation process GR1 expression was altered.
Initial Northern blot analysis (fig. 3.10) suggested that down regulation was not rapid, the mRNA still being present after 4

Figure 3.11. Kinetics of GR1 regulation on differentiation of PCC3 cells.

A. Northern blot analysis. 10ug of total cellular RNA from cells treated with retinoic acid for 0-96 hours and 6-13 days were probed with oligolabelled EcoRI fragment of pGR165 (GR165) ($7x10^5$cpm/ml used) and, subsequently, oligolabelled PstI fragment of pBSBT (αtubulin) ($8x10^5$cpm/ml used). In both cases the filter was washed to 0.2xSSC, 0.1% (w/v) SDS at 55^oC and exposed for 3 hours.

B. Samples of cells from **A.** were analysed for expression of stage-specific embryonic antigen (SSEA-1) as described in section 2.26.


A
PCC3
RA
0
4
8
12
24
36
48
60
72
84
96
6d
8d
10d
13d
GR165-
28s-
tubulin-


B
SSEA-1 Positive Cells (%)
100
90
80
70
60
50
40
30
20
10
0
0
2
4
6
8
10
12
14
Exposure to Retinoic Acid (days)

days of retinoic acid treatment. To determine more precisely the timing of the down regulation of GR1 expression, total cellular RNA was prepared from PCC3 cells at various times after the addition of retinoic acid. Samples of the differentiating culture were also analysed by immunocytochemical staining for expression of the stage-specific embryo antigen (SSEA-1), a marker characteristic of embryonic stem cells (Solter and Knowles, 1978). Throughout the time course of the experiment growth medium was replaced every 48 hours (section 2.3). 10ug of RNA from each time point was used for Northern blot analysis using the oligolabelled EcoRI fragment of pGR165 and PstI fragment of pBSBT (Table 2.3) as probes. As shown in fig. 3.11a, GR1 expression was found to decrease markedly relative to αtubulin after 4-6 days of retinoic acid treatment.

The first morphological changes in the culture were observed 24 hours after the addition of retinoic acid with the appearance of flattened cells with a high cytoplasmic: nuclear ratio (data not shown) and the culture was confluent after 72 hours of retinoic acid treatment. Immunocytochemical staining of the differentiating culture indicated the loss of SSEA-1 between 1-4 days of retinoic acid treatment (fig 3.11b). Thus, the down regulation of GR1 in the differentiating PCC3 culture is a relatively late event, occurring after the loss of the EC cell phenotype, the decrease in cell proliferation and the first signs of altered cell morphology.

Section 3.2.9. Expression of GR1 in ES cells and mouse embryogenesis.

In order to determine whether GR1 expression was regulated in the ES cell system, Northern blot analysis was performed on total cellular RNA from two ES cell lines, HD14 (XX) and CC1 (XY) (Bradley et al., 1984). RNA from both ES cell lines, and cell types generated on in vitro differentiation of each cell line, were probed with oligolabelled EcoRI fragment of pGR165 and PstI fragment of pHPT4 (Table 2.3) as a control. As shown in fig. 3.12a, no sex-linked expression of the 6.8kb GR1 transcript was observed, it being detected in both male and female ES cell lines. In both cell lines, GR1 expression was seen to increase markedly after 14 or 15 days of in vitro differentiation, the time at which the embryoid bodies were replated (section 1.2.2). After 32 days of differentiation of the HD14 cell line, GR1 expression was seen to decrease to a level similar to that seen in the undifferentiated cell type. HPRT RNA levels were seen to be approximately constant at all stages of differentiation.

The transient peak in the level of GR1 RNA corresponded to a stage at which the cystic embryoid bodies formed in suspension culture contained a high proportion of parietal and visceral endoderm cell types. After 15 days of differentiation the embryoid bodies were replated such that after 32 days of culture a variety of terminally differentiated cells had formed (Lovell-Badge, pers. comm.).

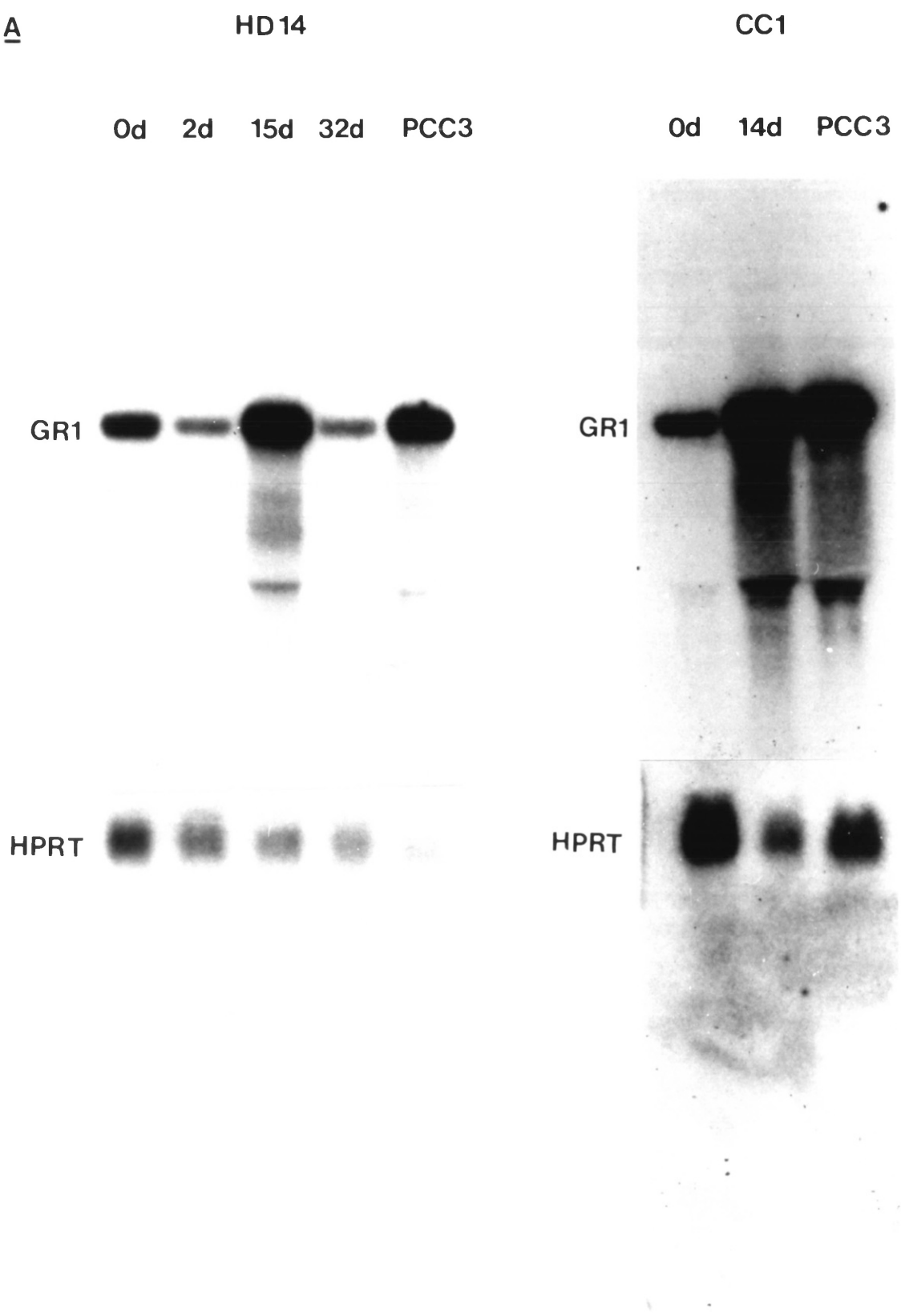
Figure 3.12. Northern blot analysis of GR1 expression in ES cells and embryogenesis.

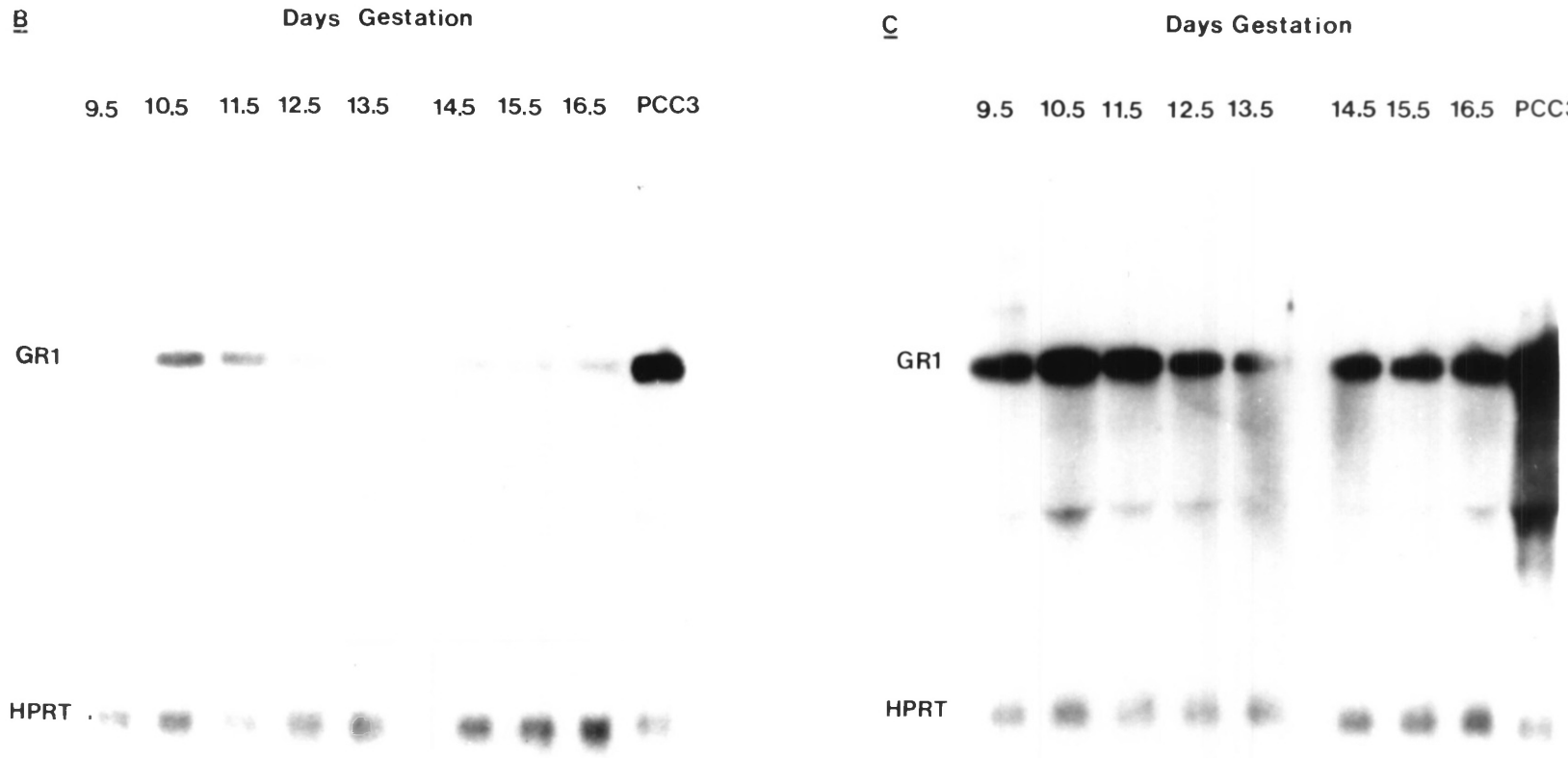
A. 10ug of total cellular RNA from ES cell lines HD14 and CC1 differentiated *in vitro* for 0, 2, 15 or 32 days (HD14) and 0 or 14 days (CC1), along with 10ug of total cytoplasmic RNA from PCC3 cells, were probed with oligolabelled EcoRI fragment of pGR165 (GR1) (10^6cpm/ml used) and oligolabelled PstI fragment of pHPT4 (HPRT) (10^6cpm/ml used). In both cases filters were washed to 0.2xSSC, 0.1% (w/v) SDS at 55^oC and exposed for 16 hours.

B. (overleaf). 10ug of total cellular RNA from mouse embryos at 9.5-16.5 days gestation and total cytoplasmic RNA from PCC3 cells were probed with oligolabelled EcoRI fragment of pGR165 (GR1) ($1.4x10^6$cpm/ml used) and oligolabelled PstI fragment of pHPT4 (HPRT) ($2.4x10^6$cpm/ml used). In both cases filters were washed to 0.2xSSC, 0.1% (w/v) SDS and exposed for 1 hour (GR1) or 6 hours (HPRT).

C. (overleaf). As for **B.** but GR1 hybridisation exposed for 16 hours.

[ES and embryo RNA samples kindly provided by Drs. Ian Jackson and Robin Lovell-Badge.]


A
HD14
CC1
0d
2d
15d
32d
PCC3
0d
14d
PCC3
GR1
GR1
HPRT
HPRT


B
Days Gestation
9.5 10.5 11.5 12.5 13.5 14.5 15.5 16.5 PCC3
GR1
HPRT
C
Days Gestation
9.5 10.5 11.5 12.5 13.5 14.5 15.5 16.5 PCC3
GR1
HPRT

In order to determine whether GR1 was expressed during mouse embryogenesis, Northern blot analysis was performed on total cellular RNA isolated from whole mouse embryos, dissected free of extraembryonic tissues, at various stages of gestation using oligolabelled EcoRI fragment of pGR165 and PstI fragment of pHPT4. As seen in fig. 3.12b,c the 6.8kb GR1 transcript was detected at all stages of embryogenesis from 9.5-16.5 days gestation. Relative to the levels of HPRT transcript, a transient peak in GR1 expression was observed at 10.5-11.5 days gestation. At all stages, the GR1 signal was markedly lower than that seen for PCC3 cells.

A relatively long autoradiographic exposure of the embryo time course Northern blot revealed the presence of two GR1 homologous transcripts, of higher molecular weight than the 6.8kb mRNA, detectable only at 9.5 days gestation (fig. 3.12c). It is unclear whether such transcripts represent unspliced precursors of the major 6.8kb GR1 mRNA, alternatively spliced transcripts of the same locus, or transcripts of other transcription units related to the GR1 locus (section 4.2).

Section 3.2.10. GR1 expression in embryonic, neonatal and adult mouse tissues.

Given the expression of GR1 at all stages of embryogenesis analysed (section 3.2.9) it was of considerable importance to

determine whether this gene showed any clear regulation *in vivo* in a manner analogous to that seen *in vitro* in the PCC3 EC cell system. In order to answer this question, total cellular RNA was prepared from a variety of tissues of neonatal (1 day old) and adult mice. Northern blot analysis of these RNA samples, along with total cellular RNA prepared from mid-gestation embryonic tissues, was performed using oligolabelled *Eco*RI fragment of pGR165 and *Pst*I fragment of pBSBT, or strand-specific radiolabelled RNA synthesised from pBS1200A or pBSBT, as probes.

As shown in figs. 3.13b and fig. 3.14, whilst the 6.8kb GR1 transcript was easily detected in the PCC3 RNA control lanes no detectable GR1 expression was found in heart, lung, liver, spleen, or kidney of newborn or adult mice, even on prolonged autoradiographic exposure. However, low levels of the GR1 transcript were seen in testis of 4 week or 9 week old mice and in adult brain. The αtubulin probe indicated that intact RNA was present in all lanes with the exception of neonatal brain RNA in which some degradation of the tubulin mRNA was observed. This degradation may account for the faint smear seen in neonatal brain with the GR1 probe.

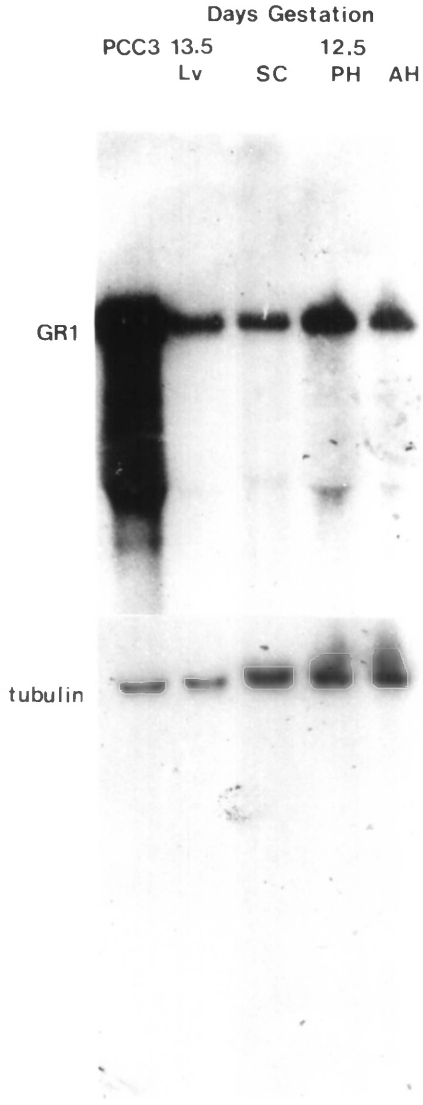
In contrast to the pattern seen in newborn and adult tissues, GR1 expression was readily detectable in RNA prepared from 12.5 days gestation spinal cord, posterior head and anterior head, and liver from 13.5 days gestation embryos (fig. 3.13a). Thus, clear developmental and tissue-specific expression of GR1 was observed during mouse development.

Figure 3.13. Northern blot analysis of GR1 expression in embryonic, newborn, and adult tissues.

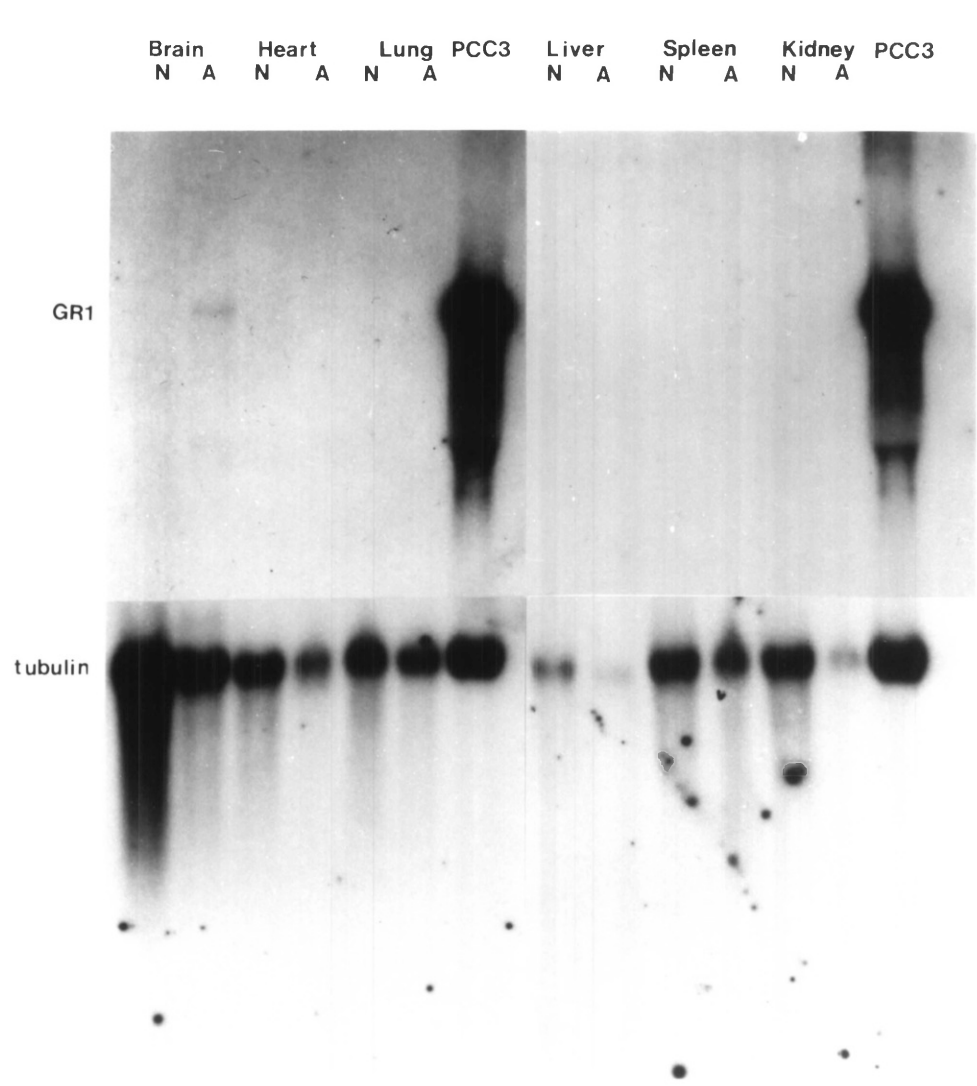
A. 10ug of total cellular RNA from 13.5 day gestation liver (Lv), 12.5 day gestation spinal cord (SC), posterior head (PH), anterior head (AH), and 10ug of total cytoplasmic RNA from PCC3 cells were probed with oligolabelled EcoRI fragment of pGR165 (GR1) ($6x10^5$cpm/ml used) and oligolabelled PstI fragment of pBSBT (tubulin) ($6x10^5$cpm/ml used). In both cases the filter was washed to 0.2xSSC, 0.1% (w/v) SDS at 55^oC and exposed for 6 hours (tubulin) or 24 hours (GR1).

[Embryonic tissues were kindly provided by Drs. Ian Jackson and Robin Lovell-Badge].

B. 10ug of total cellular RNA isolated from a number of tissues of newborn (N) or adult (A) Balb/c mice were probed with radiolabelled RNA synthesised *in vitro* using T3 polymerase on a template of pBS1200A linearised with SacI (GR1) (specific activity $1.4x10^9$cpm/ug; 10^6cpm/ml used) and pBSBT linearised with BamHI (tubulin) (specific activity $2.9x10^8$cpm/ug; 10^6cpm/ml used). In both cases filters were washed to 0.1xSSC, 0.1% (w/v) SDS at 65^oC and exposed for 8 days.


Days Gestation
PCC3 13.5
Lv
SC
12.5
PH
AH
GR1
tubulin

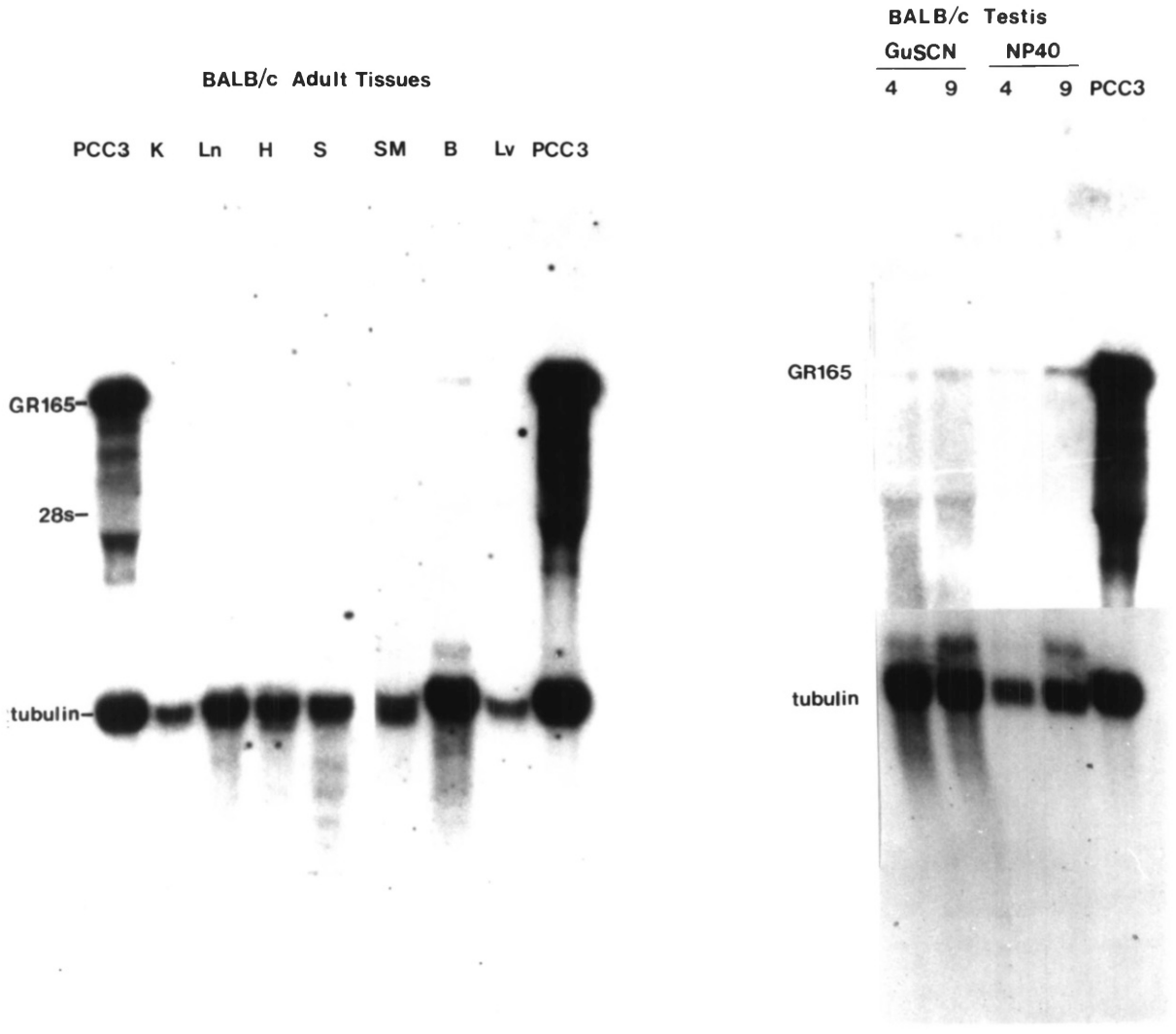
A


Brain
N A
Heart
N A
Lung
N A
PCC3
Liver
N A
Spleen
N A
Kidney
N A
PCC3
GR1
tubulin

B

Figure 3.14. Northern blot analysis of GR1 expression in adult mouse tissues.

Filters containing 10ug of total cellular RNA from adult kidney (K), lung (Ln), heart (H), spleen (S), skeletal muscle (SM), brain (B), liver (Lv), and 4 weeks (4) or 9 weeks (9) postnatal testis (both total cellular - GuSCN (section 2.9) and cytoplasmic - NP40 (section 2.10)), along with 10ug of total cytoplasmic PCC3 RNA, were probed with oligolabelled EcoRI fragment of pGR165 (GR165) ($6x10^5$cpm/ml used) and oligolabelled PstI fragment of pBSBT (tubulin) ($6x10^5$cpm/ml used). In both cases filters were washed to 0.2xSSC, 0.1% (w/v) SDS at 55^oC and exposed for 48 hours (GR165 on kidney-liver and testis filters) or 15 hours (tubulin).


BALB/c Adult Tissues
PCC3 K Ln H S SM B Lv PCC3
GR165
28s
tubulin
BALB/c Testis
GuSCN NP40
4 9 4 9 PCC3
GR165
tubulin

Section 3.3. Discussion.

Section 3.3.1. Characterisation of 1.6U.

The rationale for the isolation of 1.6U was based on the observation of Murphy et al. (1983) of qualitatively differential expression of polyadenylated RNA species homologous to a B2 repetitive element on differentiation of ES and EC cells, as well as during mouse embryogenesis. In contrast to this observation, the B2 containing cDNA clone 1.6U corresponds to transcripts which show no regulation at the level of steady-state RNA in any EC cell line tested. Also, transcriptional run on analyses in nuclei isolated from PCC3 cells, and their differentiated derivatives, or F9 cells, and F9 cells differentiated to parietal endoderm, showed high levels of constitutive expression of the 1.6U transcription unit(s) in all 4 cell types (A.R. Cooper, pers. comm.). Thus, no data were obtained to suggest 1.6U was differentially regulated between any of the cell types analysed.

The presence of two transcripts in PCC3 cells homologous to the non-B2 region of p1.6 suggests that this cDNA clone may be an artefact consisting of copies of two distinct RNA species ligated together on construction of the cDNA library. However, since neither transcript is regulated on EC cell differentiation the expression pattern of the p1.6 clone is in apparent contradiction to the data of Murphy et al. (1983). The orientation of the B2 element in the p1.6 clone provides a

possible explanation for the data.

It has been observed that only the upper strand of a B2 element is present in polysomal, cytoplasmic, polyadenylated RNA and that in such cases the B2 element is always located at the 3′ end of the mRNA (Ryskov et al., 1983). Both strands of B2 elements are detected in nuclear RNA suggesting that only the upper strand can survive RNA processing (Ryskov et al., 1983). The transcriptional orientation of the p1.6 clone, together with the nucleotide sequence determined for the B2 element, indicate that the B2 element in p1.6 is located at the 5′ end of the cDNA clone and that it is the lower strand of the element which is represented in the 1.6U mRNA. Thus, it is likely that the clone p1.6 represents an unspliced precursor transcript, the B2 element lying in an intron of the 1.6U locus. Such a potential precursor RNA species was detected in Northern blot analysis using total cellular RNA (fig. 3.3) but not in polyadenylated RNA (fig. 3.2) of PCC3 cells.

Regardless of the nature of the p1.6 cDNA clone, it is clear that whilst 1.6U is a low copy gene of novel DNA sequence, it is of little use for studies of transcriptional regulation in embryonic stem cells. However, the Northern blot analyses along with the results of transcriptional run on analyses in PCC3 and F9 cells (A.R. Cooper, pers. comm.) indicate that the 1.6U transcription unit is constitutively active at a high level in both embryonic stem cells and differentiated cell types. In this respect the cis-regulatory elements of the 1.6U

locus may be of particular use for obtaining high levels of expression of a chosen structural gene throughout development in experiments concerned with the study of gene function *in vivo*. In order to isolate these *cis*-regulatory elements, genomic clones corresponding to the 1.6U locus were isolated (section 4.2.7).

Section 3.3.2. Characterisation of λ22.

Northern blot analyses indicated that whilst expression of this gene is down regulated on *in vitro* differentiation of three EC cell lines, λ22 is also expressed in at least one SV40-transformed murine fibroblast cell line. The apparent increase in λ22 RNA levels in SV40-transformed fibroblasts, relative to normal mouse fibroblasts, should be treated with caution given the differential hybridisation signal seen with the HPRT "control" probe (fig. 3.6). Nevertheless, this result raises the possibility that at least some of the down regulation of λ22 on differentiation of EC cells may be attributable to the loss of the transformed phenotype and/or decrease in proliferation rate of such cells. Additional Northern blot analyses on a wider variety of transformed cell lines are required in order to establish any correlation of λ22 expression with the transformed phenotype.

λ22 regulation in EC cells may be similar to that of c-*myc*, the expression of which is markedly down regulated as early as

8 hours after the addition of retinoic acid to F9 cells. However, such regulation was also seen when retinoic acid treated cells were grown in the presence of butyrate, a known inhibitor of F9 differentiation (Dean et al., 1986). Similar experiments, along with Northern blot analysis of λ22 expression in EC cells in various growth states, could be carried out in order to investigate further the regulated expression of λ22 in response to changes in cell proliferation.

Regardless of the question of whether expression of λ22 responds to changes in cell proliferation or loss of embryonic or transformed phenotype of EC cells, other data suggest that this cDNA clone is not suitable for further analysis concerning the mechanism of transcriptional regulation and gene function during mouse embryogenesis.

λ22 homologous sequences are known to be present in multiple copies in the mouse genome (Gorman et al., 1985a) and transcriptional run-on analysis in nuclei from PCC3 cells indicates the 22 RNA to be transcribed from weak promoter elements (A.R. Cooper, pers. comm.). Whilst the DNA sequence determined for λ22 indicates it to be a novel gene, the moderately repetitive nature and transcript size of λ22 is reminiscent of endogenous retroviral elements in the mouse genome.

The steady-state levels of intracisternal A particle (IAP) RNA have been shown to be down-regulated on retinoic acid induced differentiation of PCC3 cells (Hojman-Montes de Oca et al.,

1983) whilst a differential screening protocol for genes activated on differentiation of F9 cells lead to the isolation of cDNA clones corresponding to IAPs (Howe and Overton, 1986). Differential screening of a cDNA library prepared from SV40-transformed fibroblasts led to the isolation of cDNA clones corresponding to the VL30 family of moderately repetitive elements, representing genes activated on SV40 transformation (Singh et al., 1985). Differential screening of a cDNA library prepared from PCC3 cells resulted in the isolation of a novel class of endogenous retroviral elements, the early transposon (ETn) elements, expressed in PCC3 and F9 EC cells but not trophoblast cells (Brulet et al., 1983).

DNA sequence comparisons of the published sequences of IAP (Burt et al., 1984), VL-30 (Singh et al., 1985) and ETn (Sonigo et al., 1987) elements with that determined for λ22 failed to reveal any homology, suggesting that this cDNA clone may represent a novel class of endogenous retroviral elements, or other type of gene family, in the mouse genome. Confirmation that λ22 represents an endogenous retrovirus could come from restriction enzyme mapping, DNA sequence analysis and Southern blotting experiments of λ22 genomic clones in a manner analogous to that used to characterise a repetitive element associated with the Ren-2 gene in the mouse (Burt et al., 1984) or the ETn elements (Kaghad et al., 1985).

Section 3.3.3. Expression of GR4.

The absence of any clear differential expression of GR4 on differentiation of any of the 3 EC cell lines tested, the expression in normal and transformed fibroblasts, combined with the low level of transcriptional activity of the GR4 transcription unit, as measured by nuclear run-on assays (A.R. Cooper, pers. comm.) preclude any further analysis of this gene with regard to the study of transcriptional regulation in embryonic stem cells.

Section 3.3.4. Regulated expression of GR1.

The Northern blot analyses carried out with the cDNA clone pGR165 clearly indicate that GR1 is worthy of further investigation concerning the study of gene regulation in murine EC cells, ES cells and mouse embryogenesis.

The expression pattern in various EC cell lines, along with the absence of detectable GR1 mRNA in normal or SV40-transformed fibroblasts, illustrate that the repression of GR1 expression seen on differentiation of PCC3 cells is not associated with the loss of the transformed phenotype. Nor is GR1 expression restricted to embryonic stem cells, activity being maintained in SSEA-1 negative cultures of differentiating PCC3 cells, F9 cells differentiated with retinoic acid and cyclic AMP, and increased on differentiation

of PCC4 and ES cell lines.

Whilst differentiation of PCC4 cells generates a homogeneous but ill-defined cell type (Jetten et al., 1979; data not shown), treatment of F9 cells with retinoic acid and cyclic AMP generates a parietal endoderm cell type (Strickland et al., 1980). The expression of GR1 in endoderm cells provides a possible explanation for the transient increase in the steady-state level of GR1 observed on differentiation of PCC3 cells and ES cell lines. The cystic embryoid bodies present after 15 days of suspension culture of ES cells contain a high proportion of parietal and visceral endoderm cell types (R. Lovell-Badge, pers. comm.). Similarly, on the basis of immunocytochemical staining data, it was concluded that the retinoic acid-induced differentiation of PCC3 cells to fibroblasts involved the transient appearance of endoderm-like cells, present in a high proportion after 5 days of differentiation (Watson, 1986). Northern blot analysis (data not shown) along with RNA in situ hybridisation studies (section 5.2.4), indicate high levels of GR1 mRNA in extraembryonic tissues including trophoblast, visceral endoderm and parietal endoderm.

The decrease in abundance of GR1 mRNA between 4-6 days of PCC3 differentiation occurred long after the culture had become confluent, suggesting that the regulation of GR1 expression in differentiating PCC3 cells is not related to the proliferation state. Relative to the regulated expression of other genes in the EC cell system the change in GR1 expression

is quite distinct, occurring late in the differentiation process. Down regulation of c-myc in F9 cells (Dony et al., 1985) and some novel genes in PSA cells (Levine et al., 1984) occurs relatively quickly after the addition of retinoic acid (within 24 and 12 hours respectively), whilst the decrease in p53 RNA levels is not observed until 2-3 days after the addition of retinoic acid to F9 cells (Dony et al., 1985). The induction of laminin and type IV collagen expression on differentiation of F9 cells occurs within 24-48 hours of retinoic acid addition (Wang et al., 1985) whilst c-fos and some murine homeobox genes show a transient, but relatively rapid, response to retinoic acid in F9 cells (Mason et al, 1985; Brier et al., 1986).

Moreover, the data suggest that the regulation of GR1 in PCC3 cells is not a direct response to retinoic acid, it being likely that GR1 is regulated indirectly by factors induced/ repressed by retinoic acid. Furthermore, the fact that the same chemical inducer of differentiation results in either a decrease or increase in expression of GR1, dependent on the EC cell line used, suggests that the regulation of GR1 expression is positioned well down the cascade of events involved in retinoic acid induced differentiation of EC cells. The positive or negative regulation of GR1 is likely to be a function of the differentiated cell type produced rather than a general feature of the effect of retinoic acid. Taken together, such data indicate that, whilst not stem cell specific, GR1 expression shows distinct cell type specificity

and its regulation may be of developmental significance.

This conclusion is strengthened by Northern blot analysis of various mouse tissues. Whilst the 6.8kb GR1 mRNA was detected at all stages of embryogenesis analysed, the level of expression was consistently lower than that seen in PCC3 cells. This indicates that either there is a generally lower level of GR1 activity in all cells of 9.5-16.5 days gestation embryos, or only a proportion of cells in the embryo express this gene. Similarly, the transient peak in GR1 RNA levels at 10.5-11.5 days gestation may be attributed to a higher proportion of GR1 expressing cells at this stage in development. Such a transient peak in gene expression has been observed for Hox-1.1 and Hox-1.2 around 12 days gestation (Colberg-Poley et al., 1985b) and En-1 between 10.5-12.5 days gestation (Joyner et al., 1985). Although Northern blot analysis of dissected mid-gestation embryonic tissues provided no evidence for tissue-specific expression, the resolution of such analyses is poor. In order to determine more precisely the expression pattern of GR1 during mouse embryogenesis, a series of RNA in situ hybridisation experiments was initiated (Chapter 5).

Whilst GR1 mRNA is readily detectable in mid gestation embryonic tissues, and whole embryos up to 16.5 days gestation, no GR1 transcripts were detected in a number of neonatal or adult mouse tissues. However, since only a limited number of tissues were analysed it remains possible that high levels of GR1 activity are maintained in some tissues not

analysed. What is clear is that, at least at a gross level, GR1 expression is regulated in a variety of tissues during mouse development. The significance, if any, of the relatively low levels of GR1 mRNA in adult brain and testis is currently unclear.

Expression of a number of other developmentally regulated genes has also been observed in adult brain and/or testis. Hox-1.1 expression is seen in adult brain and testis (Colberg-Poley et al., 1985b). A specific Hox-1.4 transcript was detected in adult testis (Duboule et al., 1986) and expression of this gene has been localised to the spermatogenic cells of adult mice (Rubin et al., 1986). The proto-oncogene int-1 has also been found to be expressed in adult testis (Jakobovits et al., 1986).

Possible roles for GR1 expression in spermatogenesis could be investigated further by Northern blot analysis of testes from prepubertal mice representing different stages of the first round of spermatogenesis, or RNA from testes of mutant mouse strains in which spermatogenesis is blocked at different stages, as carried out for other genes expressed during mammalian spermatogenesis (Willison and Ashworth, 1987). Additional localisation of GR1 expression in testis could be achieved by Northern blot analysis of RNA from purified spermatogenic cells (Wolgemuth et al., 1984) and RNA in situ hybridisation experiments to sections of testis tissue (section 5.2.4). Such studies would establish whether GR1 expression is associated with the germ cell or somatic cell

components of this organ and at what stage in spermatogenesis GR1 is expressed. The brain-specific expression of GR1 could be studied further by means of RNA *in situ* hybridisation experiments.

In a number of the Northern blot experiments a smear of GR1 homologous transcripts smaller than the characteristic 6.8kb mRNA were observed. It is unclear whether such transcripts represent alternatively initiated/ terminated or spliced transcripts of the single GR1 locus, transcripts of other loci homologous to GR1, or degradation products of the 6.8kb GR1 mRNA. A cytosine-rich repetitive element located in the GR1 cDNA clone (section 4.2.2) cannot account for these additional transcripts since they were observed with probes derived from both the entire cDNA clone and probes lacking this repetitive element. The two transcripts larger than the 6.8kb mRNA observed in RNA isolated from 9.5 days gestation embryos may represent unspliced precursors of the GR1 mRNA or transcripts of other loci containing the cytosine-rich genomic repeat. Genomic loci with homology to the cytosine-rich repeat or with weak general homology to the GR1 cDNA clone have been isolated (sections 4.2.4 and 4.2.6). Moreover, cDNA clones have been isolated from an 8.5 days gestation mouse cDNA library (provided by Dr. B.L.M. Hogan) which have strong or weak homology (in a 2:1 ratio) with the unique GR1 cDNA probe (data not shown). Further analysis of such genomic and cDNA clones will determine the nature of the additional GR1 transcripts seen on Northern blot analyses.

In conclusion, the expression pattern of GR1, in conjunction with the finding that GR1 is a low copy gene (section 4.2.1), of novel nucleotide and amino acid sequence (section 6.2), and the fact that there is a transcriptional component to GR1 regulation in differentiating PCC3 cells (A.R. Cooper, pers. comm.), make this gene suitable for further analysis with the aim of studying the mechanism of transcriptional regulation of this gene in murine development.

Chapter 4. Cloning and preliminary characterisation of the GR1 locus.

Section 4.1. Introduction.

The Northern blot analyses presented in chapter 3 indicated that one of the cDNA clones isolated in the differential library screening procedure, pGR165, was of interest for the study of the mechanisms of transcriptional regulation in murine embryonic stem cells. An essential feature of such mechanistic studies is the availability of the cis-acting regulatory elements of the GR1 transcription unit (chapter 1). Given the high level of expression of the 1.6U gene in a variety of cell types (section 3.2.1) it was also considered desirable to isolate the cis-regulatory elements of this gene. The characterisation of such regulatory elements requires the availability of genomic clones of the locus of interest. For novel genes such as GR1 and 1.6U, this is most easily achieved by nucleic acid hybridisation screening of mouse genomic libraries representing the entire genome. Having mapped the transcription unit(s) within the genomic clones, analyses can be carried out to determine the DNA sequences responsible for the regulated expression of the gene of interest.

Section 4.1.1. Cloning eukaryotic genomes in Escherichia coli.

The ability to isolate and clone unique genomic loci from a complex genome, such as that of a mammal, has been a central feature in the rapid accumulation of knowledge concerning the organisation and regulation of eukaryotic genes over the last decade.

The relatively small size of the genomes of lower eukaryotes such as Drosophila melanogaster and Saccharomyces cerevisiae facilitated the generation of representative genomic clone banks, or libraries, in plasmid vectors propagated in E.coli (Wensink et al., 1974; Carbon et al., 1977). However, the generation of the first library representative of a mammalian genome (Maniatis et al., 1978) was made possible largely as a result of technical advances concerning the generation of bacteriophage lambda cloning vectors and the use of efficient in vitro packaging extracts for the generation of viable recombinant bacteriophage particles. Sensitive screening methods using radiolabelled nucleic acid probes (Benton and Davis, 1977) allowed the isolation of clones homologous to the rabbit globin locus from this library.

Whilst bacteriophage lambda vectors were central to the generation of the first mammalian genomic libraries there is a limit to the size of genomic DNA which can be represented in a single clone. This limit (15-20kb) is a consequence of the organisation of the genes essential for the lambda life cycle and the restricted size range of DNA molecules that can be

efficiently packaged into bacteriophage particles. In order to circumvent this size restriction, cosmid vectors were developed (Collins and Hohn, 1978). These vectors consist of bacteriophage lambda cos sites, necessary for circularisation of the bacteriophage genome on infection of a host E.coli cell, along with a plasmid origin of replication and an antibiotic resistance marker gene. As such, they allow the cloning of 50kb fragments of genomic DNA, so reducing the number of clones which, statistically, represent a complete genome (Clarke and Carbon, 1976).

The use of both bacteriophage lambda and cosmid vectors has ensured that the cloning of many genomic loci is now a relatively routine task. The random nature of the clones in a genomic library allows the isolation of overlapping clones for a particular locus. Such chromosome walks have resulted in the isolation of large contiguous stretches of genomic DNA (Bender et al., 1983; Steinmetz et al., 1986). Moreover, major projects are currently underway to construct physical maps of the entire genomes of Caenorhabditis elegans and Saccharomyces cerevisiae from the restriction maps of overlapping cosmid or bacteriophage lambda genomic clones (Coulson et al., 1986; Olsen et al., 1986).

In the mammalian genome, a gap exists between the resolution of cytogenetic techniques for gene mapping (5000-10000kb) and the distances covered by conventional cosmid "walking" (300-600kb). This gap can now be filled, at least to some extent, by the use of "jumping" cosmid libraries (Collins and

Weissman, 1984) and the recently developed yeast artificial chromosome (YAC) vectors (Burke et al., 1987). Such technologies, along with the recently published human genetic linkage map (Donis-Keller et al., 1987) now makes the generation of physical maps of mammalian genomes an attainable goal.

Section 4.1.2. Instability of cloned genomic sequences.

Whilst it is apparent that the great majority of genomic loci can be cloned in E.coli using either cosmid or lambda vectors, it is known that the construction of genomic libraries does not clone sequences in a statistically random fashion (Coulson et al., 1986; Cross and Little, 1986). Differences in the relative growth rates of cosmid clones are observed in an unamplified library and such differences would be expected to lead to under-representation of certain clones on library amplification. Moreover, there are numerous anecdotal reports of the inability to isolate clones corresponding to specific genomic loci (see below). These problems, in addition to questioning the feasibility of cloning an entire eukaryotic genome, can present considerable problems in the analysis of a specific gene of interest.

Examples of genomic loci found to be unstable in conventional E.coli host strains include the human α-globin cluster (Lauer et al., 1980), a repetitive element in the genome of Kangaroo

rat (Liu and Lark, 1982), the S_u region of the mouse Ig locus (Nikaido et al., 1981), human Ig L light chain locus (Taub et al., 1983), a translocation between chromosomes 6 and 10 in a mouse plasmacytoma (Perlumter et al., 1984), actin genes of Physarum (Nader et al., 1985), the D14S1 polymorphic locus of human (Wyman et al., 1985) and clones associated with the Prader-Willi syndrome on human chromosome 15 (Donlon et al., 1986).

A number of strategies have been employed to determine the influence of both vector and host E.coli genotypes and/or phenotypes on the representative recovery of genomic libraries. Most of these studies have been concerned with the use of bacteriophage lambda vectors. Some of these factors are illustrated in figure 4.1. and discussed below.

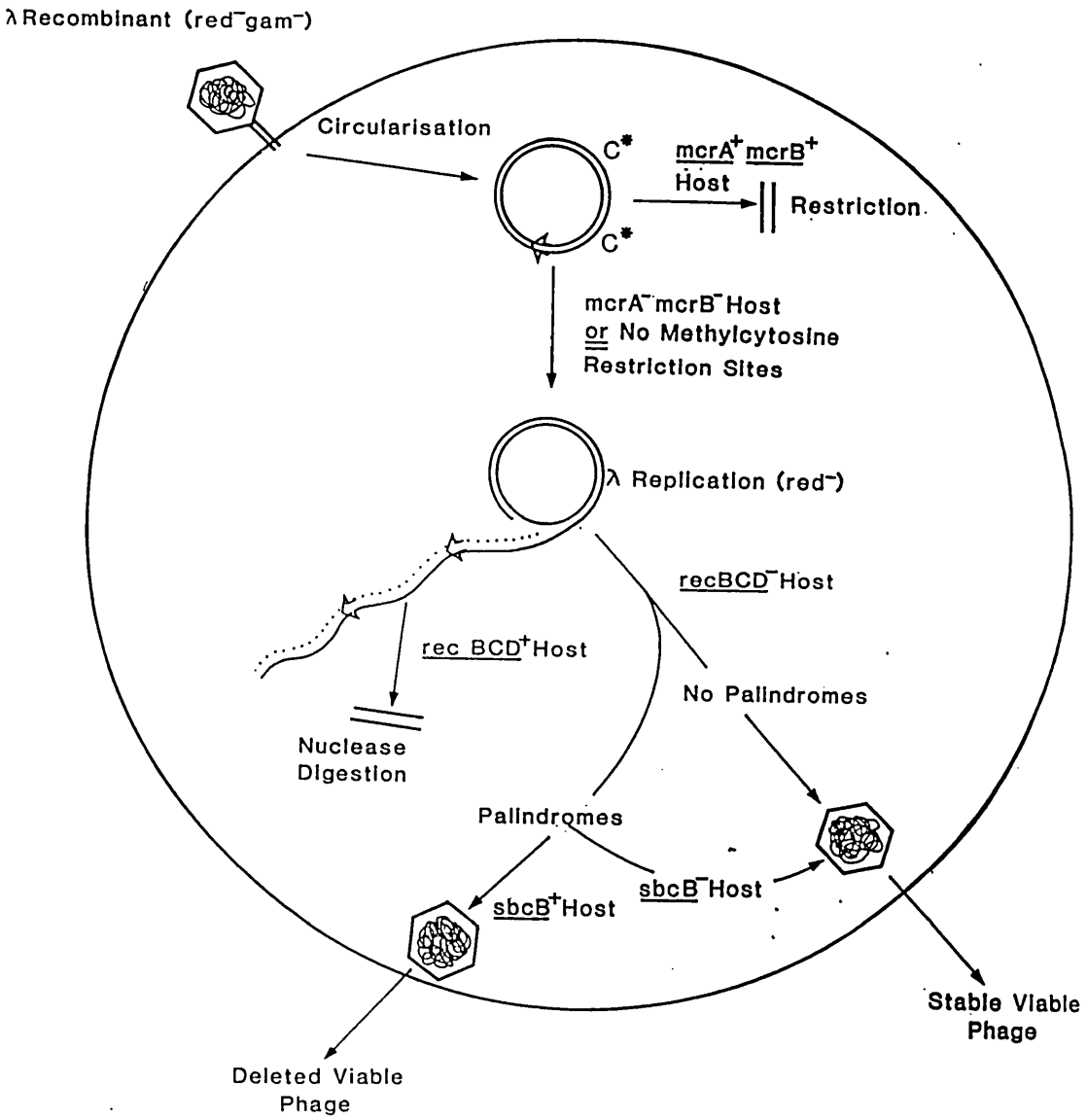
Section 4.1.3. Factors influencing the propagation of recombinant genomic clones.

(i) Recombination systems.

Bacteriophage lambda vectors used to construct genomic libraries are generally replacement vectors in which the non-essential region of the vector genome is replaced with eukaryotic genomic DNA in a recombinant library (Brammar, 1982). The non-essential region usually contains the gam and red genes which play a role in the normal life cycle of bacteriophage lambda.

Figure 4.1. Potential problems in cloning specific mammalian genomic loci.

The roles of host E.coli McrA and McrB restriction systems and recBCD and sbcB loci in the initial propagation of a red$^-$ gam$^-$ recombinant lambda clone containing mammalian genomic DNA are illustrated (see section 4.1.3). The possible fates of the recombinant clone are shown. C*= target sequences in mammalian genomic DNA for the Mcr systems. => = lambda cos site.


λ Recombinant (red⁻gam⁻)
Circularisation
C*
C*
mcrA⁺ mcrB⁺
Host
Restriction
mcrA⁻ mcrB⁻ Host
or No Methylcytosine
Restriction Sites
λ Replication (red⁻)
recBCD⁻ Host
rec BCD⁺ Host
No Palindromes
Nuclease
Digestion
Palindromes
sbcB⁻ Host
sbcB⁺ Host
Stable Viable
Phage
Deleted Viable
Phage

The gam gene product antagonises the activity of exonuclease V, the product of the recBC locus of E.coli (Enquist and Skalka, 1973; Friedman and Hays, 1986), so allowing efficient lambda replication via the rolling circle mechanism. Recently, an additional E.coli gene, recD, has been shown to encode a component of the recBC complex (Chadhury and Smith, 1984; Amundsen et al., 1986). recD is thought to encode the nucleolytic activity of the recBCD complex.

In the absence of gam, lambda replication can proceed by means of recombination of monomeric circles to generate multimers from which linear monomers can be packaged. This latter mode of replication, in the absence of red, is dependent on the general recombination system of E.coli ($RecA^+$) and the presence of a crossover hotspot instigator (chi) site in the lambda vector (Enquist and Skalka, 1973; Lam et al., 1974; Wyman et al., 1986). Given that many of the eukaryotic loci known to be unstable as lambda clones in E.coli contain repetitive elements or represent highly polymorphic loci, and undergo rearrangements on propagation, it is desirable to maintain these clones under non-recombinogenic conditions. Thus, when screening a red^- gam^- genomic library, a $recBC^-$ or $recD^-$ E.coli strain would be the host of choice. Alternatively, replacement vectors retaining gam, such as Charon 32 (Loenen and Blattner, 1983) or host E.coli strains bearing a regulated gam gene in a plasmid vector (Crouse, 1985) could be used.

Given the desire to propagate the library under non-

recombinogenic conditions it might be thought that recA⁻ host E.coli strains would be useful for library screening. However, recA⁻ recBC⁻ E.coli strains grow very poorly and, as such, are of little use in the screening of recombinant libraries.

Host E.coli exonuclease I, the product of the sbcB locus, has been shown to play a role in the stability of extrachromosomal DNA in E.coli. Collins et al. (1982) showed that the deletion of palindromic sequences in the E.coli transposon Tn5 was markedly reduced when grown in a recBC⁻ sbcB⁻ host strain. It was subsequently found (Leach and Stahl, 1983) that lambda molecules bearing synthetic palindromes were viable only when grown in a recBC⁻ sbcB⁻ host. That such a host E.coli strain can facilitate the cloning of eukaryotic genomic loci has been shown with the actin genes of Physarum (Nader et al., 1985), many human genomic clones (including the D14S1 locus) which could not be propagated on conventional host strains (Wyman et al., 1985), and the cloning of regions of human chromosome 15 rearranged in individuals with Prader-Willi syndrome (Donlon et al., 1986).

There is also at least one case where mutation in recE has facilitated the propagation of a lambda genomic clone. A highly repetitive dispersed element in the genome of the Kangaroo rat, whilst stable in plasmid vectors, underwent deletions when propagated in a lambda vector. This element was stabilised either when cloned in a red⁻ vector or grown in a recE⁻ host strain (Liu and Lark, 1982).

(ii) Methylcytosine restriction.

A proportion of cytosine residues in mammalian genomes are known to be methylated at ^{5}C to produce 5 methylcytosine (5meC), generally found as meCpG dinucleotides (Ehrlich and Wang, 1981). Maintenance of this methylation state is achieved by the activity of DNA methyltransferases (Sneider and Potter, 1969) of which there are no equivalents in prokaryotes.

Two methylcytosine restriction systems, McrA and McrB, have been found to be active in E.coli K12 strains (Raleigh and Wilson, 1986). The Mcr system is identical to the rgl system described by Revel (1967). The McrB restriction system is believed to cleave DNA at a specific sequence which includes the consensus GpmeC (Raleigh and Wilson, 1986). The restriction site of McrA is not known but does include meC. Thus, the Mcr system of E.coli has the potential to inhibit the cloning of certain mammalian loci which contain a high proportion of GpmeCpG residues (fig. 4.1).

Methylation of CpG residues of DNA in vitro, using cloned murine modification methylase, confers sensitivity to the Mcr system (Blumenthal, 1987). A role for the McrB locus has been implicated in the reduced frequency of transformation of E.coli by pBR322 DNA following propagation of the plasmid in mouse cells (Hanahan et al., 1980; Raleigh et al., 1988). Moreover, recovery of the murine Prn-p locus was achieved only when an unamplified genomic library was propagated on an McrB^{-} host E.coli strain (Westaway et al., 1987).

Clearly, the Mcr system will inhibit only a proportion of

genomic clones containing GpmeCpG sequences. The greater the GpmeCpG content of the genomic sequence, the greater the probability of restriction by the Mcr system. The absence of an equivalent methylation system in E.coli ensures that after the initial rounds of lambda replication the molecule will no longer be susceptible to such restriction. However, if the locus of interest is susceptible to Mcr restriction, it is essential to screen unamplified libraries.

Whilst there are numerous potential problems in cloning a specific locus, it should be emphasised that the majority of loci are stable in E.coli. Thus, when attempting to isolate genomic clones for a novel gene, considerations such as those given above should be employed only after screening libraries on conventional host strains has failed to detect the locus of interest.

Having established a restriction map of the GR1 locus, attempts were made to isolate GR1 genomic clones from amplified cosmid and lambda mouse genomic libraries. This resulted in the isolation of three cosmid clones containing a moderately repetitive element found in the cDNA clone pGR165. Clones corresponding to the GR1 locus were not recovered.

A genomic library from PCC3 genomic DNA was constructed in the bacteriophage lambda vector lambda fix and the unamplified library screened on recD$^-$ McrA$^-$ McrB$^-$ and recBC$^-$ sbcB$^-$ host E.coli strains for GR1 homologous sequences. The library was also screened for 1.6U homologous clones. In both cases, the

appropriate genomic clones were isolated only from the library plated on the Mcr deficient host strain.

Restriction map and Southern blot analyses of the pGR165 homologous genomic clones confirmed that one, N28, corresponded to the GR1 locus and contains 18kb of DNA transcriptionally upstream from the GR1 cDNA clone. The other clones appear to represent a locus (or loci) with weak homology to pGR165.

Section 4.2. Results.

Section 4.2.1. Restriction map of GR1 locus.

Initial Southern blot analyses of mouse genomic DNA, using the entire cDNA insert of pGR165, indicated that GR1 was a low copy gene (data not shown). In order to characterise this locus at the molecular level, a series of overlapping restriction enzyme fragments from the cDNA clone pGR165 (fig. 4.2) were used as radiolabelled probes in Southern blot analyses of restriction fragments of mouse genomic DNA.

Restriction maps of the GR1 locus in the murine EC cell line PCC3 and the murine SV40-transformed fibroblast cell line SV3T3 C138 were found to be identical (fig. 4.3). Moreover, the data indicated that the cDNA clone pGR165 mapped directly onto the GR1 locus, there being no detectable intervening sequences in the region of the GR1 locus represented by pGR165. Thus, it would appear that pGR165 represents a single exon of the GR1 transcription unit, at least 2.5kb in size.

One exception to the colinearity of the restriction maps of pGR165 and the corresponding region of the GR1 locus was found. Two StuI sites within the region of homology with pGR165 were detected in genomic DNA on the basis of Southern blots using the 1200ASacT3 probe of pGR165 (fig. 4.2 and fig. 4.9). The cDNA clone pGR165 contained only one of these restriction enzyme sites. This apparent discrepancy could be indicative of the presence of a small intron in the GR1 locus

Figure 4.2. cDNA probes used for mapping GR1 genomic locus.
The restriction fragments of pGR165 indicated were used to generate radiolabelled DNA probes by the method of random priming or radiolabelled RNA probes (section 2.17) in the case of 1200ASacT3.
E= EcoRI, Hc= HincII, K= KpnI, P= PstI, S= SacI, St= StuI. The black area designates the position of an element found to be moderately repetitive in the mouse genome (section 4.2.2).

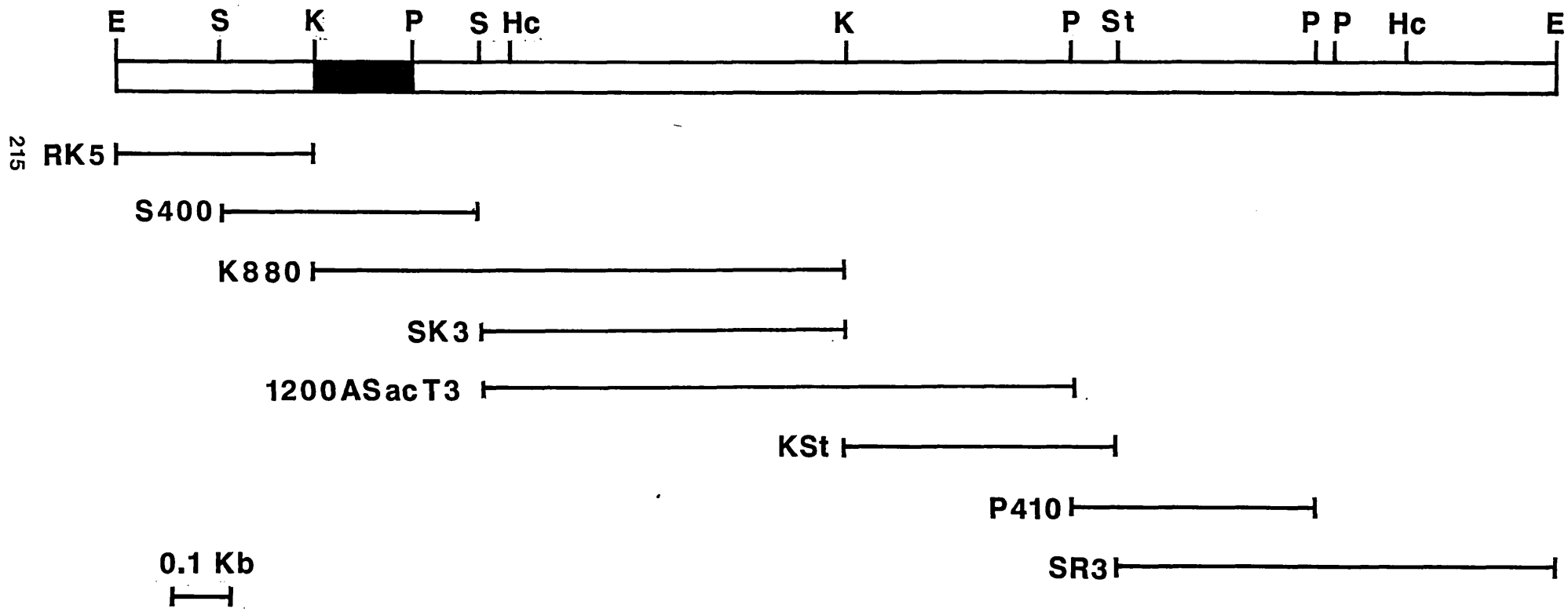
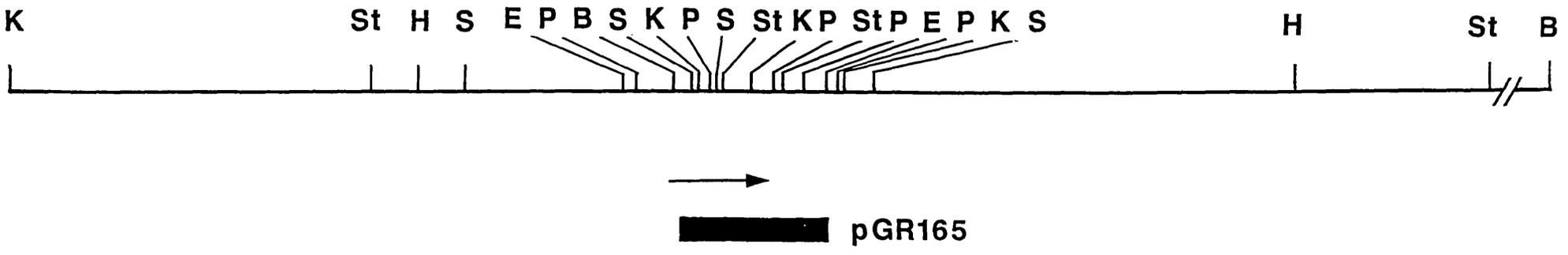

E
S
K
P
S
Hc
K
P
St
P
P
Hc
E
RK5
S400
K880
SK3
1200ASacT3
KSt
P410
SR3
0.1 Kb

Figure 4.3. Restriction map of GR1 locus.

The restriction fragments of pGR165 given in fig. 4.2 were used as probes in Southern blot analyses of PCC3 or SV3T3 C138 genomic DNA to give the restriction map of the GR1 locus. The position of the cDNA clone pGR165 relative to this locus is shown by the black box. The arrow designates the direction of transcription of the GR1 gene (section 5.2.1). B= BamHI, E= EcoRI, H= HindIII, K= KpnI, P= PstI, S= SacI, St= StuI. The 3′ BamHI site was estimated to lie ~17Kb downstream from the 3′ end of the region represented by pGR165.



1Kb

separating two exons represented in pGR165. Alternatively, this observation could be a consequence of the activity of dcm methylase in E.coli at the StuI restriction site (see later).

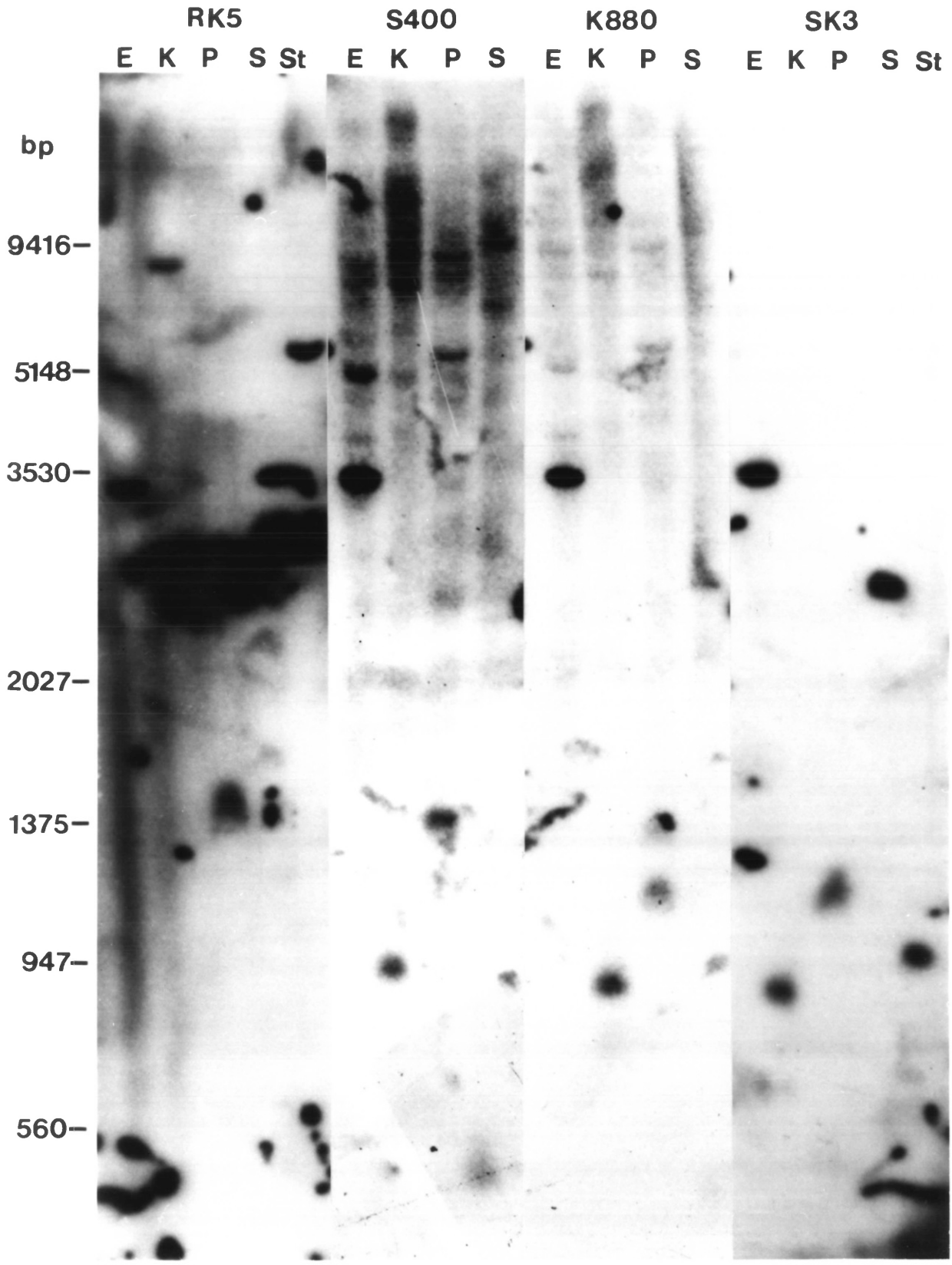
Section 4.2.2. A moderately repetitive element within the cDNA clone pGR165.

During the construction of the restriction enzyme site map of the GR1 locus (fig. 4.3) it was found that two of the restriction fragments of pGR165 were homologous to numerous restriction enzyme fragments of mouse genomic DNA. As shown in fig. 4.4, on a relatively long exposure, pGR165 restriction fragments RK5 and SK3 hybridised only to unique restriction fragments in SV3T3 C138 genomic DNA, compatible with the restriction map of the GR1 locus (fig. 4.3).

In contrast, pGR165 restriction fragments S400 and K880 hybridised not only to genomic fragments characteristic of the GR1 locus, such as the 3.5kb EcoRI restriction fragment seen with all four probes, but were also found to be homologous to a number of additional restriction fragments of genomic DNA (fig. 4.4). This result could not be attributed to incomplete digestion of genomic DNA since the same restriction digest mixture was analysed with all 4 probe fragments of pGR165. Additional analysis revealed that the the 1.1kb PstI-PstI fragment of pGR165 hybridised only to unique genomic sequences (data not shown). Thus, the moderately repetitive element was

Figure 4.4. Southern blot analysis of a repetitive element in pGR165.

High molecular weight genomic DNA isolated from SV3T3 C138 cells was digested with the restriction enzymes indicated and 10ug samples electrophoresed through a 0.7% (w/v) agarose gel. Following transfer to nitrocellulose as described in section 2.18, each filter was probed with one of RK5, S400, K880 or SK3 restriction enzyme fragments of pGR165 (fig. 4.2), radiolabelled by the method of random priming (section 2.17.1 (specific activities $9.6x10^7$cpm/ug, $6.7x10^7$cpm/ug, $7.13x10^7$cpm/ug and $7.9x10^7$cpm/ug respectively; approximately 10^6cpm/ml used for each hybridisation). Filters were washed to 0.2xSSC 0.1% (w/v) SDS at 65^oC and exposed for 16 hours (S400 and K880 filters) or 3 days (RK5 and SK3 filters). The size of hybridising fragments was estimated by comparison with the migration of bacteriophage lambda DNA digested with HindIII or EcoRI and HindIII. bp= base pairs. E= EcoRI, K= KpnI, P= PstI, S= SacI, St= StuI.


RK5
S400
K880
SK3
E K P S St
E K P S
E K P S
E K P S St
bp
9416–
5148–
3530–
2027–
1375–
947–
560–

localised to a 170bp KpnI-PstI fragment within the cDNA clone pGR165.

Section 4.2.3. Screening amplified mouse libraries for GR1 genomic clones.

Having established that the cDNA clone pGR165 represents the transcription unit of a low copy gene, attempts were made to isolate genomic clones corresponding to the GR1 locus. Three different restriction fragments of pGR165 were used as radiolabelled hybridisation probes to screen three amplified cosmid, and one amplified lambda, mouse genomic libraries (fig. 4.5). All of these libraries had previously been used successfully in the isolation of clones corresponding to a number of loci.

The C3H spleen genomic library in the cosmid vector pTCF (Grosveld et al., 1982) had been used to isolate non-rearranged cosmid clones corresponding to neurofilament, H-2K, H-2D and numerous overlapping MHC clones, whilst neurofilament and numerous MHC clones (but not H-2D^d) had been previously isolated from the K36 cosmid library (F. Grosveld, pers. comm.). The mouse/human chromosome 11 hybrid genomic library in the cosmid vector LoristB (Cross and Little, 1986) contained the predicted proportion of human clones (2%), and had previously been used for the isolation of non-rearranged clones for human catalase and numerous randomly picked human

clones. A number of mouse type II collagen cosmid clones had also been isolated previously from this library (P.F.R. Little and K.S.E. Cheah, pers. comm.). The DBA2/J liver genomic library in the bacteriophage lambda vector L47 (Loenen and Brammar, 1980) had been used to isolate clones corresponding to murine Ren-1 and Ren-2 loci (Mullins et al., 1982) and other clones flanking these loci (W.J. Brammar, pers. comm.). Using the calculation of Clarke and Carbon (1976), it was estimated that it was necessary to screen more than $2x10^5$ cosmid clones or $6x10^5$ bacteriophage plaques from each library to ensure a probability >99% of detecting a single clone of a single copy locus. This calculation assumes that all genomic loci are equally represented in the library. An appropriate number of clones from each library were plated, duplicate filters prepared (section 2.18), and hybridised with radiolabelled restriction fragments of pGR165 as indicated in figure 4.5.

Despite the number of clones screened, and the previous success in cloning other loci from these libraries, only from the K36 cosmid library were pGR165 homologous genomic clones isolated. These 3 clones, CAD18, CAD20 and CAD22, were purified by further rounds of bacterial colony hybridisations and DNA prepared from each for restriction map and Southern blot analyses.

Figure 4.5. Screening amplified genomic libraries for GR1 genomic clones.

Various radiolabelled restriction fragments of pGR165 (fig. 4.2) were used as hybridisation probes to screen the mouse genomic libraries indicated. The C3H spleen and K36 cell line libraries constructed in the cosmid vector pTCF (Grosveld et al., 1982) were the kind gift of K.M. Hui and F. Grosveld. The cosmid library constructed from the mouse/human hybrid cell line IB5 in the cosmid vector LoristB (Cross and Little, 1986) was kindly provided by P.F.R. Little. The genomic library constructed from DBA2/J liver in the bacteriophage lambda vector L47 (Mullins et al., 1982) was the kind gift of W.J. Brammar. Genotypes of E.coli strains are given in fig. 2.1.

Figure 4.5. Screening amplified mouse libraries for GR1 genomic clones.

pGR165 probe	Genomic library	Host *E.coli* strain	No. clones screened	Positively hybridising clones
Entire cDNA	C3H spleen in pTCF	ED8767	1.2×10^6	None
K880	K36 cell line in pTCF	ED8767	2.0×10^6	CAD 18, CAD 20, CAD 22.
1200A SacT3	mouse/human hybrid in LoristB	ED8767	8.0×10^5	None
1200A SacT3	DBA2/J liver in L47	ED8910	6.0×10^5	None

Section 4.2.4. Analysis of pGR165 homologous cosmid clones.

Initial characterisation of the pGR165 homologous cosmid clones CAD18, CAD20 and CAD22 was limited to the construction, for each clone, of a partial restriction site map of the region of homology with the pGR165 cDNA.

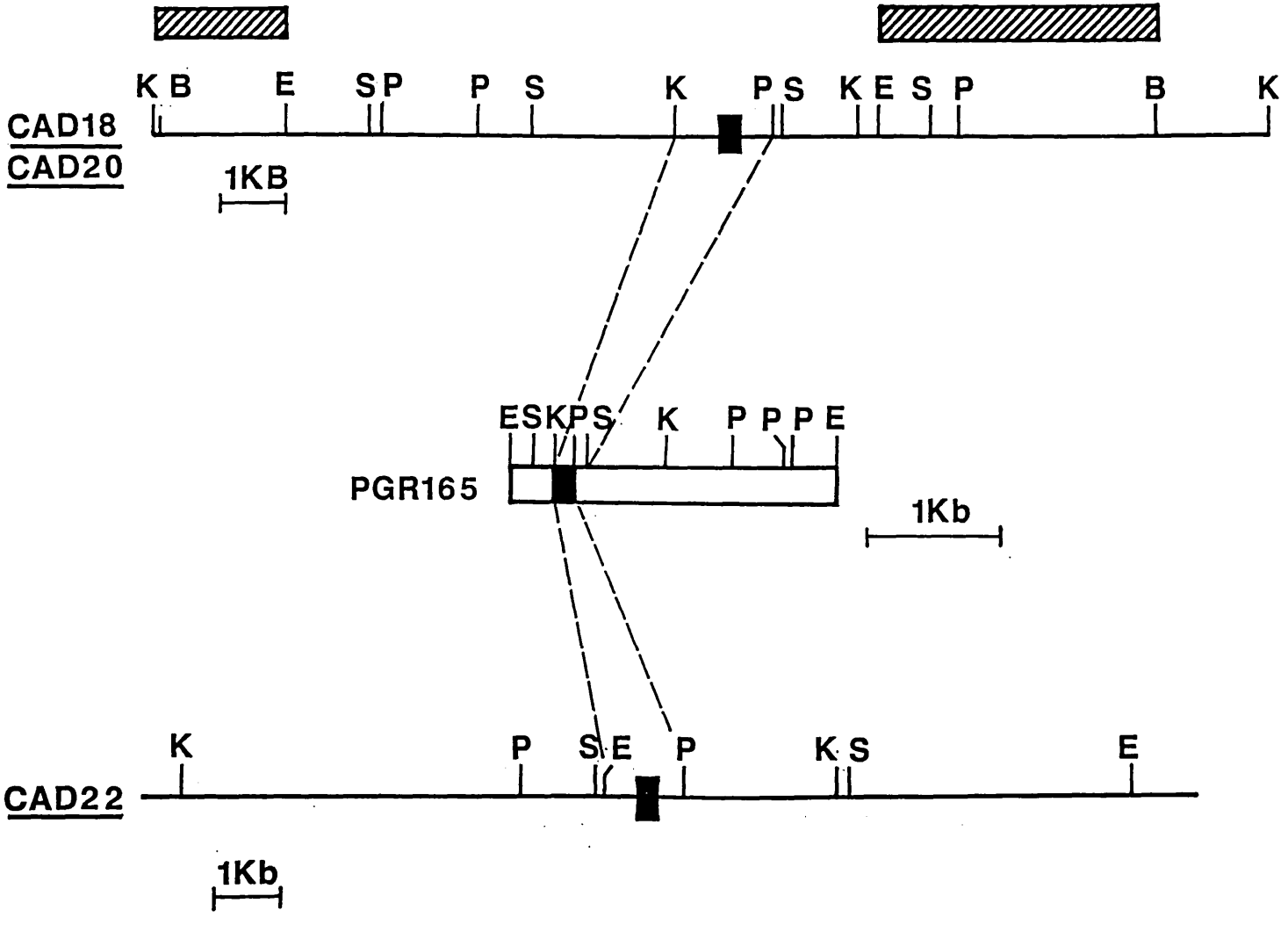
Each of the pGR165 cDNA restriction fragment probes used in the characterisation of the GR1 locus (fig. 4.2) was used in Southern blot analyses of restriction digests of CAD18, CAD20 and CAD22 (data not shown). Two restriction fragments of CAD18 and CAD20 were also used as hybridisation probes in the restriction mapping studies. The partial restriction enzyme maps of CAD18, CAD20 and CAD22 are given in figure 4.6.

The maps of CAD18 and CAD20 were identical. This, along with the cross-hybridisation observed between the two clones in regions other than that shared with pGR165 (data not shown), suggested that CAD18 and CAD20 represent the same or closely related loci which are distinct from that represented by CAD22.

Whilst all 3 cosmid clones contained sequences homologous to K880 and S400 probes of pGR165, no homology was detected with any of the other regions of pGR165. The radiolabelled 1.1kb PstI-PstI fragment of pGR165 also failed to hybridise to the cosmid clones, so localising the region of homology between CAD18/20, CAD22 and pGR165 to a 170bp KpnI-PstI fragment of the cDNA clone. Given the absence of any large intervening sequences between the coding sequences represented by pGR165

Figure 4.6. Restriction map of CAD18, CAD20 and CAD22 genomic clones.

Radiolabelled restriction fragments of the cDNA clone pGR165 were used in Southern blot analyses in order to construct limited restriction maps of cosmid clones CAD18, CAD20 and CAD22. Two BamHI-EcoRI restriction fragments of CAD18 and CAD20 (designated by the hatched boxes) were used as probes in additional Southern blot analyses. The maps of CAD18 and CAD20 were identical. The region of homology between pGR165 and CAD18, CAD20 and CAD22 is designated by the black box. The EcoRI and KpnI restriction fragments of CAD22 with homology to pGR165 were estimated to be 16kb and 15kb, respectively, in size. The scale of the pGR165 restriction map is twice that of the cosmid clones. B= BamHI, E= EcoRI, K= KpnI, P= PstI, S= SacI.


K B E S P P S K P S K E S P B K
CAD18
CAD20
1KB
E S K P S K P P P E
PGR165
1Kb
K P S E P K S E
CAD22
1Kb

(section 4.2.1) it was concluded that the cosmid clones CAD18, CAD20 and CAD22 represent mouse loci containing the moderately repetitive element in pGR165 (section 4.2.2) and are otherwise unrelated to the GR1 locus.

Thus, it was apparent that genomic loci corresponding to the GR1 locus were not present (at least at a detectable frequency) in any of the amplified genomic libraries analysed. As such, alternative strategies would have to be employed to clone the GR1 locus.

Section 4.2.5. Construction of a mouse genomic library in bacteriophage lambda.

Given the failure to isolate genomic clones corresponding to the GR1 locus (sections 4.2.3 and 4.2.4), it was evident that this locus was unstable when propagated either as a cosmid or bacteriophage lambda molecule in E.coli.

As discussed earlier (section 4.1.2), both cosmid and bacteriophage lambda genomic libraries are known not to be representative of the entire genome. Differential growth rates of specific genomic clones are known to vary and could result in under-representation of certain loci in amplified libraries. The presence of at least one copy of a CG-rich repetitive element in the GR1 locus (sections 4.2.2 and 6.2.4) suggested that propagation of this locus in E.coli might be susceptible to recombination and/or methylcytosine restriction

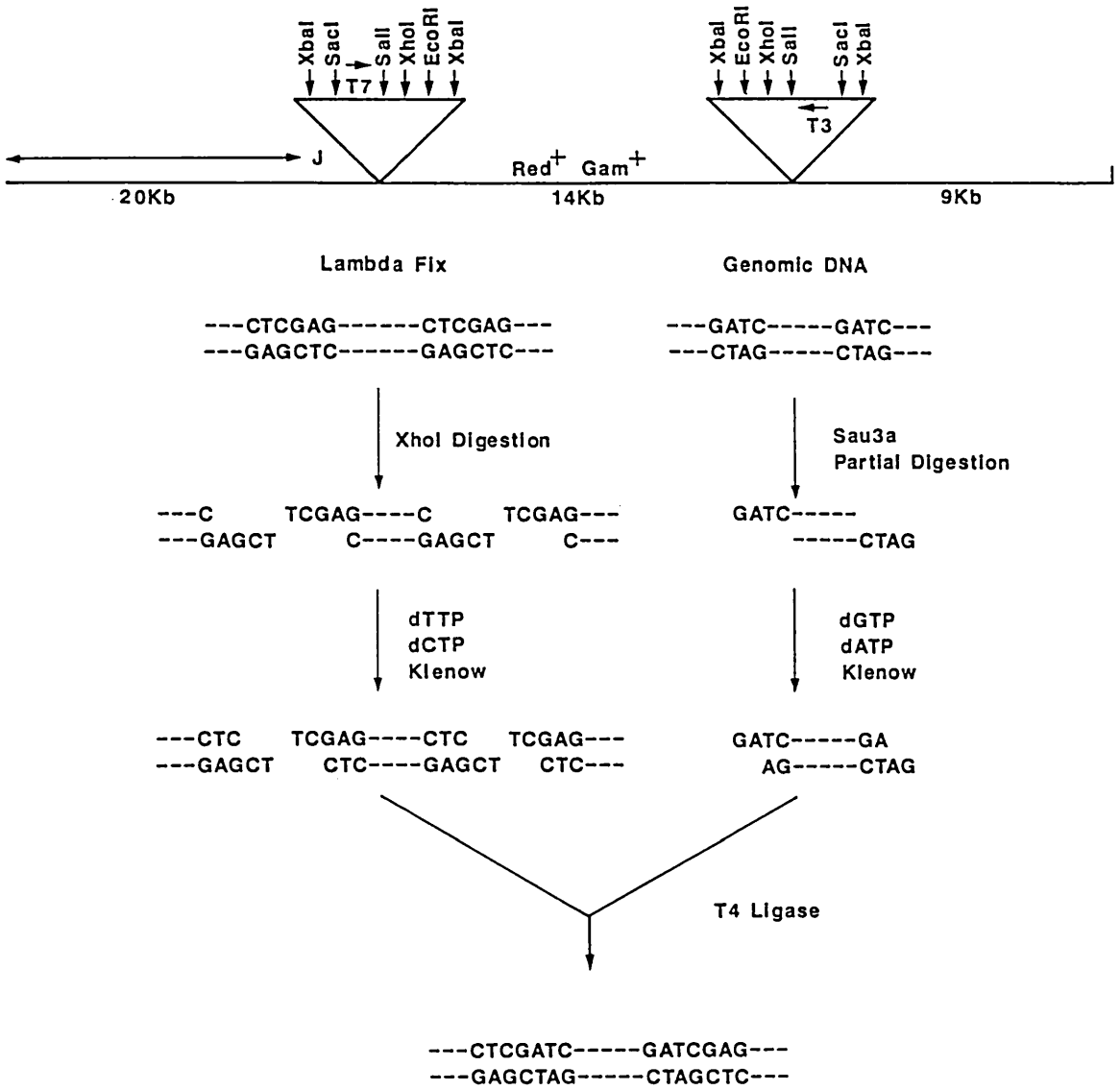
(section 4.1.3). The amplified libraries screened with pGR165 probes had been propagated on recBC$^-$ sbcB$^+$ host E.coli strains and E.coli ED8767 is known to have an McrB$^+$ phenotype (Raleigh et al., 1988).

In an attempt to overcome the problems encountered in cloning the GR1 locus, a mouse genomic genomic library was constructed from PCC3 genomic DNA in the bacteriophage lambda vector lambda fix (fig.4.7). The in vitro packaged bacteriophage particles were then plated on each of two host E.coli strains, NM621 (recD$^-$ sbcB$^+$ McrB$^-$) and CES 200 (recBC$^-$ sbcB$^-$ McrB$^+$). The first strain would facilitate the propagation of clones containing a high proportion of 5meC residues, whilst the latter strain would allow the cloning of palindromic sequences (section 4.1.3).

Stages in the construction of the genomic library are shown in figures 4.7 and 4.8. Given the limited capacity of lambda vectors (Brammar, 1982) only genomic fragments of 15-22kb would generate bacteriophage molecules of packagable size. Optimum conditions for Sau3a partial restriction of PCC3 genomic DNA to an average size of 15-22kb were established (section 2.14) and a large scale digestion carried out. Restriction fragments were separated through a sucrose gradient (section 2.16.2) and samples of fractions from this gradient subjected to electrophoresis in a 0.4% (w/v) agarose gel (fig. 4.8a). Fractions were pooled and samples of each pool digested to completion with KpnI. The restriction fragments generated were subjected to Southern blot analysis

Figure 4.7. Efficient ligation of genomic and bacteriophage vector restriction fragments.

The relevant features of the bacteriophage lambda vector lambda fix (a modification of λ2001 (Karn et al., 1984)) are shown. Some of the steps in the generation of the PCC3 genomic library in lambda fix are illustrated. Partial filling-in of XhoI digested vector DNA and Sau3a partially digested genomic DNA optimises the formation of recombinant lambda molecules containing a single insert of genomic DNA.

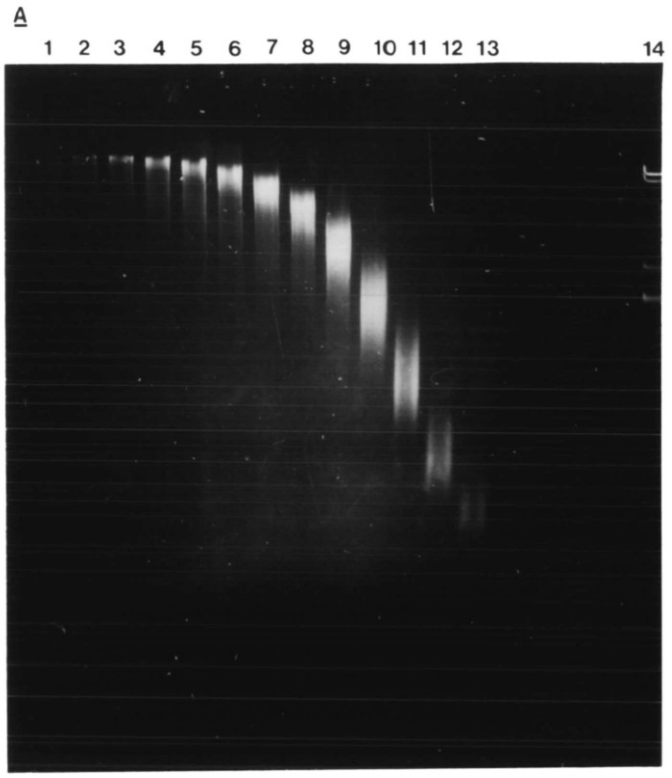

XbaI
SacI
T7
SalI
XhoI
EcoRI
XbaI
XbaI
EcoRI
XhoI
SalI
SacI
XbaI
T3
J
Red+ Gam+
20Kb
14Kb
9Kb
Lambda Fix
Genomic DNA
---CTCGAG------CTCGAG---
---GAGCTC------GAGCTC---
---GATC-----GATC---
---CTAG-----CTAG---
XhoI Digestion
Sau3a
Partial Digestion
---C TCGAG----C TCGAG---
---GAGCT C----GAGCT C---
GATC-----
-----CTAG
dTTP
dCTP
Klenow
dGTP
dATP
Klenow
---CTC TCGAG----CTC TCGAG---
---GAGCT CTC----GAGCT CTC---
GATC-----GA
AG-----CTAG
T4 Ligase
---CTCGATC-----GATCGAG---
---GAGCTAG-----CTAGCTC---

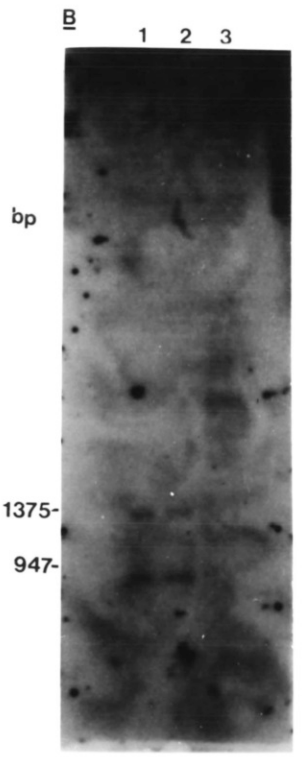
in order to establish that GR1 genomic sequences were present in the partially restricted fractions.

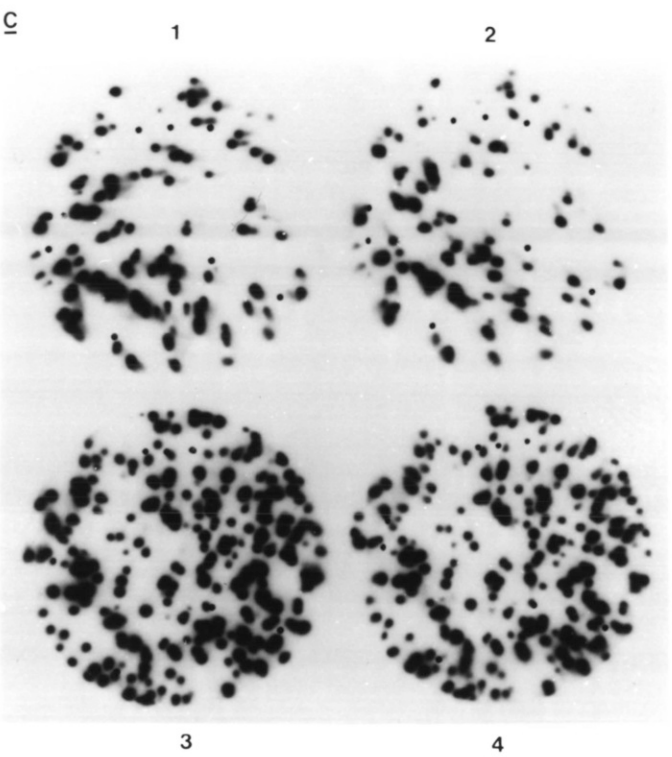
As seen in figure 4.8b, KpnI restriction fragments of 1.3kb and 900bp, characteristic of the GR1 locus, were detected on Southern blot analysis of pooled samples 6-21 and 22-28 using the non-repetitive region of pGR165 as probe. Pooled fractions 22-28 were then used for the construction of the library. The Sau3a termini of these genomic fragments were modified by the method of Zabarovsky and Allikmets (1986), ligated with similarly modified XhoI restriction fragments of lambda fix and packaged *in vitro* (section 2.16.8). The partial filling-in of both vector and genomic fragments prevents self ligation of vector and ligation of multiple genomic fragments (fig. 4.7). From the plating efficiency of the packaging mix on *E.coli* LE392 (*rec*$^+$ *sbcB*$^+$ McrB$^+$) (1.63×10^6pfu/ml) and a P2 lysogen of LE392 (1.39×10^6pfu/ml) it was estimated that 85% of the unamplified packaged library was recombinant. To confirm the presence of genomic sequences in this library, aliquots were plated on each of *E.coli* NM621 and CES 200, transferred to duplicate nitrocellulose filters, and probed with a murine B2 repetitive element (Murphy *et al.*, 1983). As shown in figure 4.8c, positively hybridising clones were detected. These were calculated as representing approximately 30% of clones, regardless of the host *E.coli* strain used. The library titre on *E.coli* NM621 (5.55×10^6pfu/ml) was found to be approximately 2 fold that seen on *E.coli* CES 200 (2.92×10^6) and 3.4 fold that seen on LE392. Such small variations in titre may be

Figure 4.8. Construction of bacteriophage lambda library from PCC3 genomic DNA.

A. Fractions of partially digested genomic DNA from the sucrose gradient were electrophoresed in a 0.4% agarose gel. 1= fraction 3, 2= fraction 6, 3= fraction 9, 4= fraction 12, 5= fraction 15, 6= fraction 18, 7= fraction 21, 8= fraction 24, 9= fraction 27, 10= fraction 30, 11= fraction 33, 12= fraction 36, 13= fraction 39, 14= bacteriophage lambda DNA digested with EcoRI and HindIII. B. DNA samples from pooled fractions 6-21 (1), 22-28 (2) and 29-42 (3) from A. were digested with KpnI, electrophoresed, fragments transferred to nitrocellulose as described in section 2.18 and probed with radiolabelled RNA synthesised in vitro using T3 polymerase on a template of pBS1200A linearised with SacI (specific activity $1.6x10^9$cpm/ug; $1.7x10^6$cpm/ml used). The filter was washed to 0.1xSSC 0.1%(w/v) SDS at 65^oC and exposed for 8 days. The size of hybridising fragments was estimated from the rate of migration of EcoRI-HindIII restriction fragments of bacteriophage lambda DNA. bp= base pairs. C. Duplicate nitrocellulose filters were prepared from 1.5ul aliquots of a 10^1 dilution of the products of in vitro packaging of the lambda fix-genomic DNA ligation mix plated on E.coli strains CES 200 (1 and 2) or NM621 (3 and 4). Filters were probed with radiolabelled RNA synthesised in vitro using SP6 polymerase on a template of p64B2 linearised with HindIII (specific activity $3.9x10^8$cpm/ug; 10^6cpm/ml used). Filters were washed to 0.1xSSC 0.1%(w/v) SDS at 65^oC and exposed for 16 hours.


A
1 2 3 4 5 6 7 8 9 10 11 12 13 14


B
1 2 3
bp
1375-
947-


C
1
2
3
4

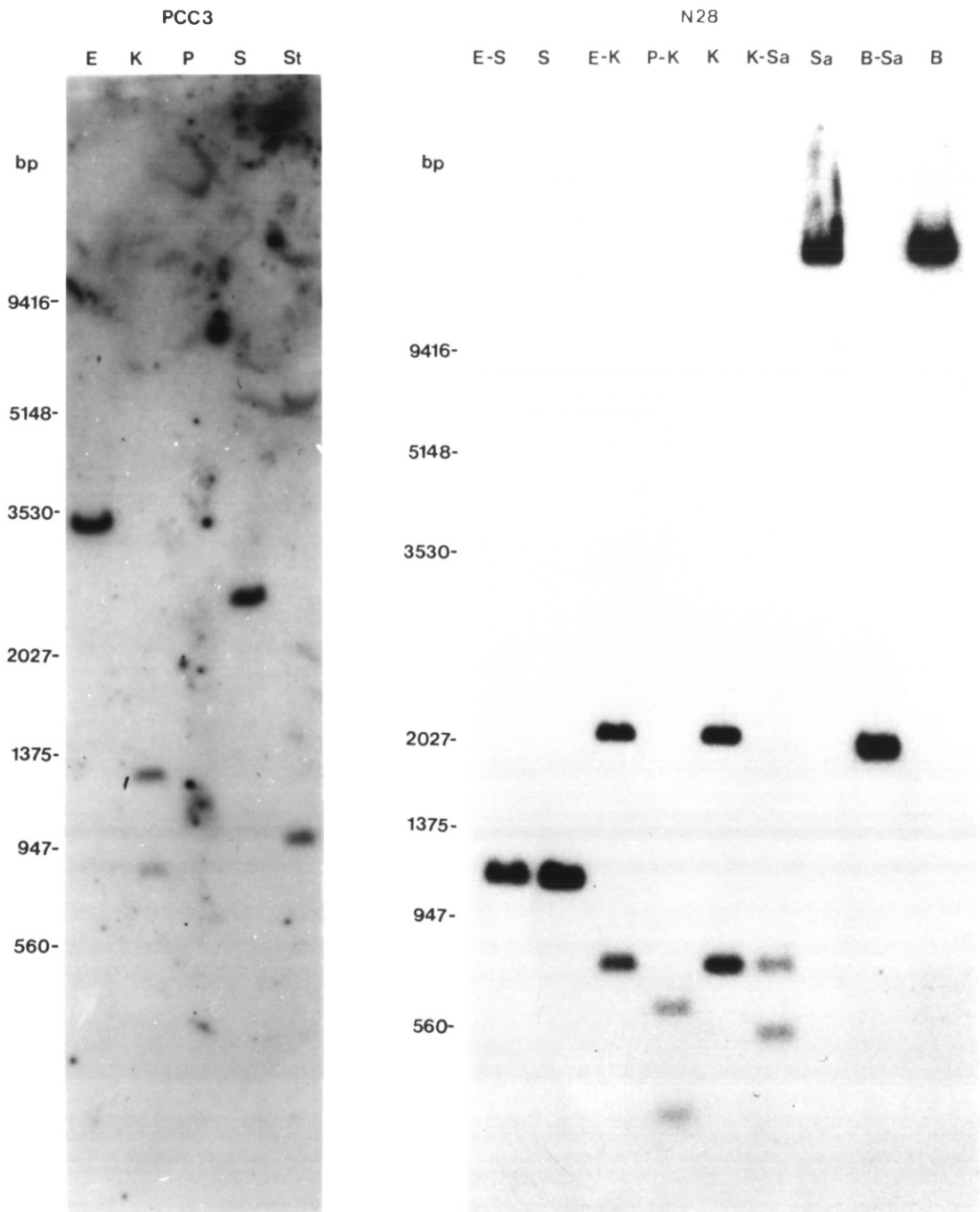
attributable to the non-isogenic nature of these strains.
Having established that the packaged bacteriophage contained a high proportion of genomic clones, the library was screened for pGR165 homologous sequences.

Section 4.2.6. Isolation of pGR165 homologous genomic clones.

$3x10^5$ p.f.u. of the unamplified PCC3 genomic library in the bacteriophage lambda vector lambda fix (section 4.2.5) were plated on each of E.coli strains CES200 (recBC$^-$ sbcB$^-$ McrB$^+$) and NM621 (recD$^-$ sbcB$^+$ McrB$^-$). Duplicate nylon membranes were prepared as described in section 2.18.4 and hybridised with radiolabelled RNA synthesised in vitro using T3 polymerase on a template of pBS1200A linearised with SacI (fig. 3.5).
Following autoradiography, one clone, N28, gave a strong reproducible hybridisation signal whilst two others, N20.1 and N20.3, gave weaker but similarly reproducible signals. All three clones were detected in the recombinant clones grown in E.coli NM621 (data not shown). Subsequent propagation was carried out in E.coli CES200. The clones were plaque purified by further rounds of hybridisation screening and DNA prepared from each as described in section 2.5.
Southern blot analysis of restriction fragments of clone N28 (fig. 4.9) indicated the presence of a 880bp KpnI restriction fragment with strong homology to the pBS1200A RNA probe. This fragment is characteristic of the GR1 locus. Additional

Figure 4.9. Southern blot analysis of N28 genomic clone.

10 ug of PCC3 genomic DNA or 1ug of N28 bacteriophage DNA was digested with the restriction enzymes indicated, the fragments electrophoresed, transferred to nitrocellulose as described in section 2.18 and probed with radiolabelled RNA synthesised *in vitro* using T3 polymerase on a template of pBS1200A linearised with SacI (specific activity 1.7×10^9cpm/ug; 10^6cpm/ml used). Filters were washed to 0.1xSSC 0.1% (w/v) SDS at 65^oC and exposed for 4 days (genomic DNA) or 3 hours (N28 DNA). B= BamHI, E= EcoRI, K= KpnI, P= PstI, S= SacI, St= StuI, Sa= SalI, E-S= EcoRI and SacI, E-K= EcoRI and KpnI, P-K= PstI and KpnI, K-Sa= KpnI and SalI, B-Sa= BamHI and SalI. The size of restriction fragments was estimated from the rate of migration of bacteriophage lambda DNA digested with HindIII or EcoRI-HindIII. bp= base pairs.


PCC3
E K P S St
bp
9416-
5148-
3530-
2027-
1375-
947-
560-
N28
E-S S E-K P-K K K-Sa Sa B-Sa B
bp
9416-
5148-
3530-
2027-
1375-
947-
560-

Southern blot experiments and restriction mapping studies confirmed that N28 represents a region of the GR1 locus. The restriction map determined for N28 (fig. 4.10) was entirely in agreement with that of the GR1 locus (fig. 4.3) indicating that this clone had not undergone rearrangement during growth in the E.coli strains used. The pGR165 homologous region of N28 is adjacent to the long arm of the lambda fix vector, there being 18kb of genomic DNA representing previously uncloned regions of the GR1 locus. In contrast to N28, Southern blot analysis of clones N20.1 and N20.3 (fig. 4.11) indicated that these clones do not represent the GR1 locus. In addition to the weak hybridisation, the EcoRI restriction fragments of these clones with homology to pGR165 were estimated to be 2.2kb in size, not the characteristic 3.5kb hybridising EcoRI fragment seen for the GR1 locus in PCC3 genomic DNA (fig. 4.9). As such, it was concluded that these clones represent a locus (or loci) related to, but not identical with, the GR1 locus.

Figure 4.10. Partial restriction map of N28 genomic clone.

A partial restriction map of the insert fragment of the genomic clone N28 was constructed from the results of Southern blot experiments using various restriction digests of N28 and radiolabelled fragments of pGR165 as probes. The region of homology with pGR165 is shown by the black box. The open boxes represent the bacteriophage lambda vector. Only the closest restriction site for a given enzyme to the GR165 homologous region is given, with the exception of SalI where a complete restriction map is presented. The arrow designates the direction of transcription of the GR1 locus (section 5.2.1). B= BamHI, E= EcoRI, H= HindIII, K= KpnI, P= PstI, S= SacI, Sa= SalI, St= StuI, Xb= XbaI.

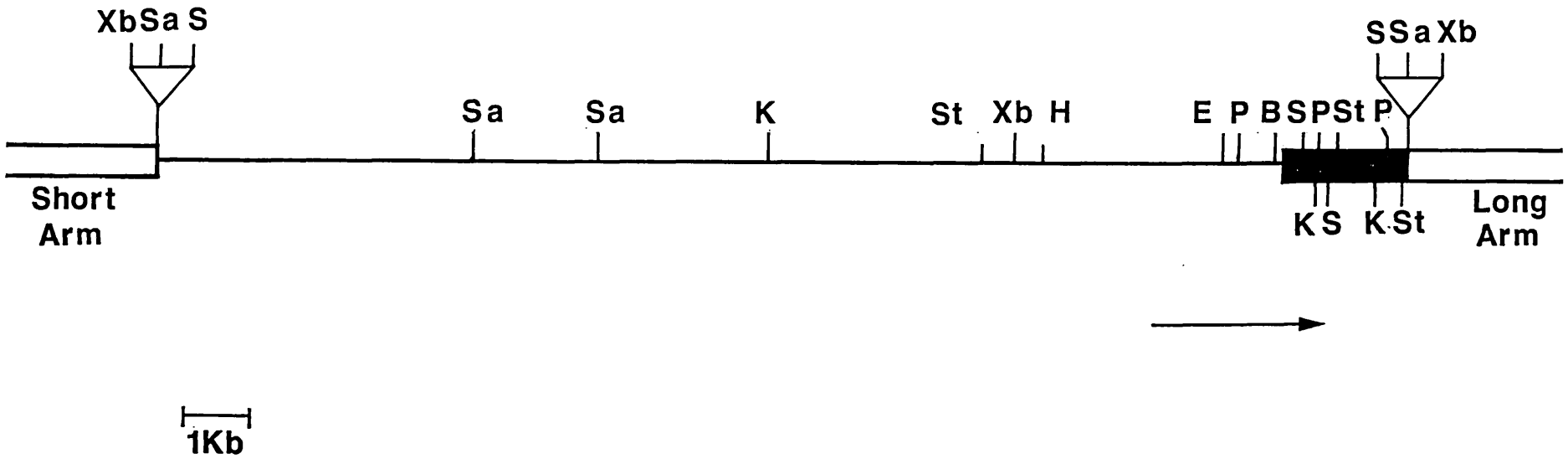
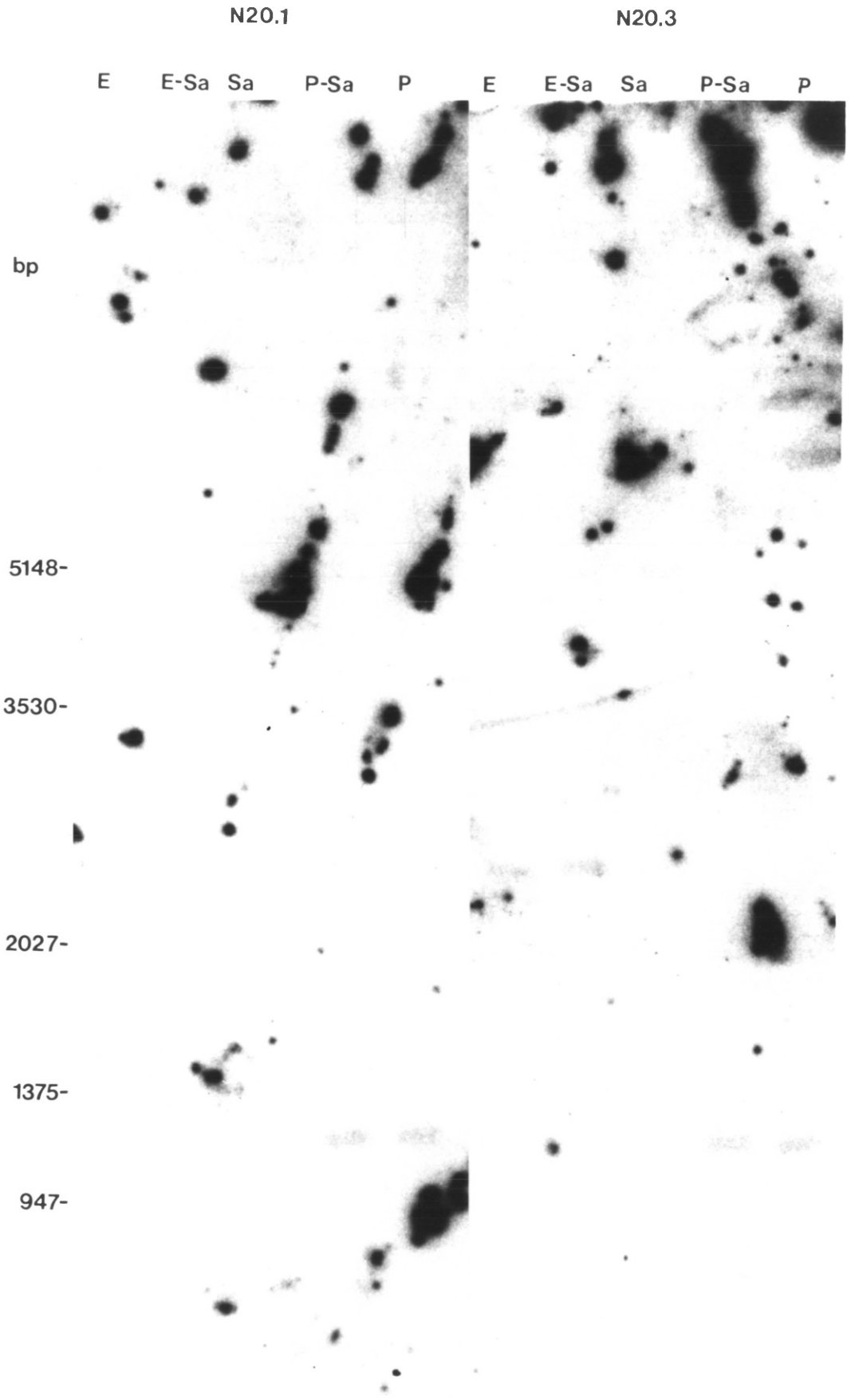

Xb Sa S
Sa
Sa
K
St
Xb
H
E
P
B
S
P
St
P
S Sa Xb
Short Arm
K S
K St
Long Arm
1Kb

Figure 4.11. Southern blot analysis of N20 genomic clones.

1ug of N20.1 and N20.3 bacteriophage DNA samples were digested with the restriction enzymes indicated, the fragments electrophoresed, transferred to nitrocellulose as described in section 2.18 and probed with radiolabelled RNA synthesised *in vitro* using T3 polymerase on a template of pBS1200A linearised with *Sac*I (specific activity 1.7×10^9cpm/ug; 10^6cpm/ml used). Filters were washed to 0.1xSSC 0.1% (w/v) SDS at 65^oC and exposed for 7 days. E= *Eco*RI, P= *Pst*I, Sa= *Sal*I, E-Sa= *Eco*RI-*Sal*I, P-Sa= *Pst*I-*Sal*I. The size of hybridising fragments was estimated from the rate of migration of *Eco*RI-*Hind*III digested bacteriophage lambda DNA. bp= base pairs.


N20.1
N20.3
E
E-Sa
Sa
P-Sa
P
E
E-Sa
Sa
P-Sa
P
bp
5148-
3530-
2027-
1375-
947-

Section 4.2.7. Isolation of 1.6U homologous genomic clones.

Screening of the amplified mouse genomic library in the bacteriophage lambda vector L47 (Loenen and Brammar, 1980) (fig. 4.5) with a radiolabelled RNA probe synthesised *in vitro* using T3 polymerase on a template of pBS1.6U linearised with *Kpn*I (fig. 3.1) failed to isolate any 1.6U homologous genomic clones (data not shown).

The same filters used for the isolation of pGR165 homologous sequences (section 4.2.6) were screened for the presence of 1.6U genomic clones using the same RNA probe as previously. From the $3x10^5$ p.f.u. plated on each of *E.coli* NM621 and CES200, five clones reproducibly hybridised with the 1.6U cDNA probe (data not shown). All five clones were isolated from library aliquots propagated on the *rec*D$^-$ McrB$^-$ host *E.coli* strain, NM621. Subsequent propagation and analysis of these clones was performed on *E.coli* CES200, as for the GR1 homologous sequences (section 4.2.6).

Southern blot analysis of each of the clones N1.1, N12.1, N13.1, N13.2, and N14.1 confirmed the presence of 1.6U homologous restriction fragments (fig. 4.12). The observation that all five 1.6U homologous genomic clones were isolated from the library grown in *E.coli* NM621 suggested that there may be features of the 1.6U locus (or loci) which inhibited propagation of 1.6U genomic clones in *E.coli* CES200. However, no such inhibition was seen during plaque purification of these clones on CES200 and four of the clones showed

equivalent plating efficiencies on each of the E.coli strains NM621, CES200 and LE392 (fig. 4.13).

Figure 4.12. Southern blot analysis of 1.6U genomic clones. 1ug samples of bacteriophage DNA from each of the 1.6U genomic clones N1.1, N12.1, N13.1, N13.2, N13.3, and N14.1 were digested with PstI (a) or PstI and SalI (b), fragments electrophoresed, transferred to nitrocellulose as described in section 2.18, and filters probed with radiolabelled RNA synthesised *in vitro* using T3 polymerase on a template of pBS1.6U linearised with KpnI (specific activity 1.25×10^9cpm/ug; 10^6cpm/ml used). Clone N13.3 was included as a negative control. Filters were washed to 0.1xSSC 0.1% (w/v) SDS at 65^{o}C and exposed for 16 hours. The size of hybridising restriction fragments was estimated from the rate of migration of bacteriophage lambda DNA digested with HindIII or EcoRI and HindIII. bp= base pairs.

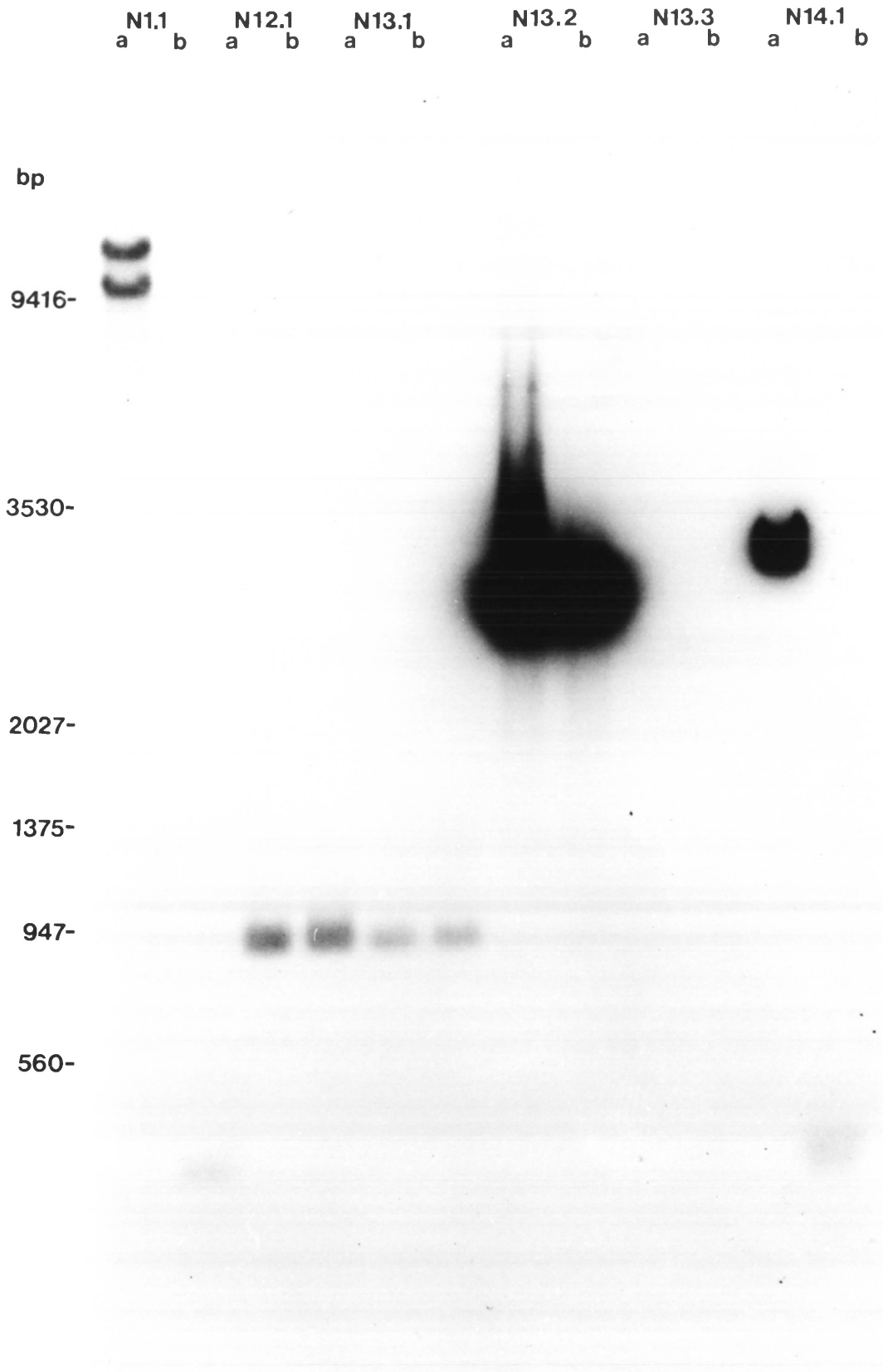

N1.1
a b
N12.1
a b
N13.1
a b
N13.2
a b
N13.3
a b
N14.1
a b
bp
9416-
3530-
2027-
1375-
947-
560-

Figure 4.13. Titre of 1.6U genomic clones isolated on NM621. The titre (p.f.u./ml) of each of the 1.6U genomic clones N1.1, N12.1, N13.2 and N14.1, originally isolated on NM621, was determined on the host E.coli strains indicated as described in section 2.2.

Figure 4.13. Titre of 1.6U genomic clones on different E.coli strains.

Clone	Host (p.f.u./ml)		
	NM621	CES200	LE392
N1.1	2.0×10^9	7.2×10^9	2.4×10^9
N12.1	1.8×10^9	4.0×10^9	8.0×10^8
N13.2	2.6×10^9	7.4×10^9	1.4×10^9
N14.1	1.8×10^9	8.2×10^9	3.4×10^9

Section 4.3. Discussion.

Section 4.3.1. Characterisation of the GR1 locus.

The use of a variety of restriction fragments from the cDNA clone pGR165 as probes for Southern blot analyses of PCC3 and SV3T3 C138 genomic DNA facilitated the construction of a restriction map of the GR1 locus.

Whilst it was clear that pGR165 represents a region of the transcription unit of a single copy gene, sequences within a 170bp KpnI-PstI restriction fragment of the cDNA clone were found to be present in numerous copies (approximately 10-20) in the mouse genome. Screening amplified genomic libraries resulted in the isolation of three cosmid clones, probably representing two loci, which contain this repetitive element. No other regions of homology with the pGR165 cDNA clone were detected in these cosmid clones.

Subsequent DNA sequence analysis established that the 170bp KpnI-PstI restriction fragment containing the repetitive element includes a cytosine and guanine rich region which, in the GR1 transcription unit, is likely to encode a proline-rich amino acid sequence (section 6.2.5). It is likely that this sequence represents the repetitive element observed on Southern blot analysis. Clearly, it would be of interest to determine whether the pGR165 homologous regions of clones CAD18/20 or CAD22 are components of a transcription unit, as well any regulated expression of such transcription units in

embryonic stem cells. Such analysis could be carried out by means of Northern blotting, SI nuclease protection assays, and DNA sequence analysis.

The restriction map determined for the GR1 locus was identical in the genomes of both GR1-expressing (PCC3) and non-expressing (SV3T3 C138) cell types. Thus, at least within the region of the GR1 locus accessible to the cDNA probes used, gene rearrangement does not appear to play a role in the regulation of GR1 expression. However, it is possible that DNA rearrangements in regions of the GR1 locus not mapped in this work may have occurred between the two cell types.

The restriction mapping studies of the GR1 locus indicated that the cDNA clone pGR165 represents a single exon of the GR1 locus of at least 2.5kb. The additional StuI restriction site detected in the GR1 locus, but not in the pGR165 cDNA clone, suggested the presence of a small intron within the region of the GR1 locus present in pGR165. However, DNA sequence analysis showed identity between the cDNA clone and the corresponding region of the GR1 locus and indicated the presence of a StuI recognition sequence in the cDNA clone in the same position as that observed in the GR1 locus (section 6.2.4). An explanation for these observations was provided by analysis of the DNA sequence flanking the predicted StuI restriction site.

The activity of StuI (recognition sequence 5'AGGCCT 3') is known to be sensitive to dcm methylation. The dcm methylase of E.coli methylates internal C residues in the sequence 5'CCTGG

3′ (May and Hattman, 1975). Thus, if the the StuI recognition sequence is followed by two guanine residues (i.e. 5′AGGCCTGG 3′) it will be protected from digestion by StuI in DNA propagated in dcm^+ host E.coli strains. This was found to be the case for the StuI restriction site in question. Mammalian cells lack an equivalent methylation system and so such a sequence in genomic DNA isolated directly from mammalian cells would be susceptible to digestion by StuI.

The EcoRI site at the 5′ end of the pGR165 was absent from the corresponding region of the GR1 locus, a predictable result given the use of synthetic EcoRI linkers in the cDNA cloning protocol used in the synthesis of pGR165 (Watson, 1986). In contrast, the EcoRI site at the 3′ end of pGR165 was found to be coincident with an EcoRI site in the GR1 locus.

It has been observed that the use of EcoRI methylase to protect internal EcoRI sites in the cDNA cloning process is often an inefficient process (F. Poirier, pers. comm.). As such, it is likely that the EcoRI site at the 3′ end of pGR165 lies within the GR1 transcription unit and is not the result of the addition of synthetic linker sequences.

This conclusion was strengthened with the finding from DNA sequence analysis that pGR165 contains no polyA sequence or consensus polyadenylation signal (section 6.2.4). These results, along with the known transcriptional orientation of pGR165 (section 5.2.1), led to the conclusion that the cDNA clone pGR165 is likely to represent the 5′ sequence of a previously longer double stranded cDNA molecule which was

cleaved at an internal EcoRI restriction site during the cDNA cloning procedure.

This conclusion could be tested by analysis of other GR1 homologous cDNA clones and genomic clones containing the 3′ region of the GR1 locus. An alternative hypothesis, that pGR165 is the result of deletion of a previously longer cDNA clone unstable on propagation in E.coli, is considered unlikely since no internal deletions within the cDNA clone were detected on comparison of restriction maps of cDNA and GR1 locus and any deletion is unlikely to have retained the 3′ EcoRI site.

The significance of this conclusion lies in the prediction that the cDNA clone pGR165 may represent sequences somewhat closer to the transcriptional start of the GR1 gene than expected if this cDNA clone were to contain the 3′ terminal sequences of this transcription unit.

Section 4.3.2. Cloning the GR1 locus.

Some considerable difficulty was encountered in the isolation of genomic clones corresponding to the GR1 locus. As outlined earlier, a small proportion of mammalian genomic loci appear to be unstable when cloned in E.coli and so will be under-represented in a genomic library. Given the negative results obtained from screening four, otherwise proven, genomic libraries for GR1 genomic clones, it was concluded that the

GR1 gene was one such unstable locus.

The presence of at least one copy of a CG-rich repetitive element in the region of the GR1 locus represented by the cDNA probe used for library screening, suggested that, like some other mammalian loci unstable in conventional host E.coli strains, it might be possible to clone the GR1 locus by propagating an unamplified mouse genomic library in a methylcytosine restriction (Mcr) deficient and exonuclease I deficient host E.coli strain. However, such a recBCD$^-$ sbcB$^-$ McrB$^-$ strain was not available (D.F.R. Leach, pers. comm.). Since it was unclear which, if either, of the Mcr or exonuclease I systems of E.coli was responsible for the under-representation of GR1 genomic clones it was decided to use both McrB$^-$ and sbcB$^-$ host strains for initial screening of the library. Since methylation of CpG residues is not maintained in E.coli, clones were subsequently grown in the sbcB$^-$ host strain.

This strategy proved successful in isolating a single clone of the GR1 locus, along with two clones corresponding to an additional locus with weak homology to the pGR165 cDNA probe. Additionally, five clones representing the 1.6U locus (or loci) were isolated. No 1.6U genomic clones had previously been isolated from the amplified mouse genomic library in the bacteriophage vector L47. Interestingly, all eight clones were isolated from aliquots of the unamplified library propagated on E.coli NM621. No GR1 or 1.6U genomic clones were detected in an equivalent number of clones of the same library

grown in E.coli CES200. All 8 clones were subsequently grown in CES200 with no appreciable decrease in titre being observed. Four of the 1.6U genomic clones were shown to plate with equivalent efficiencies on each of recD$^-$ (NM621), recBC$^-$ (CES200) and rec$^+$ (LE392) host E.coli strains.

Although not conclusive, these data, along with those of other workers, strongly suggest that the McrB restriction system can significantly reduce the representation of certain mammalian loci in recombinant libraries propagated in E.coli (Raleigh et al., 1988). Given the non-isogenic nature of CES200 and NM621, the role of factors other than the McrB system in the cloning of GR1 and 1.6U loci cannot, formally, be ruled out. The role (if any) of the sbcB mutation in preventing deletion of regions of the genomic clones was not investigated but could be answered by restriction map analysis of bacteriophage DNA of each clone grown in either sbcB$^-$ or sbcB$^+$ host E.coli strains.

Section 4.3.3. GR1 genomic clones.

Regardless of the role played by the host E.coli genotype, it is clear from restriction mapping and hybridisation studies that clone N28 represents a region of the GR1 locus. This was subsequently confirmed by DNA sequence analysis (section 6.2.3).

When the known transcriptional orientation of the cDNA clone

pGR165 (section 5.2.1) is considered, it is apparent (fig. 4.10) that N28 contains 18kb of the GR1 locus transcriptionally upstream of the cloned cDNA, the 3' end of the genomic clone likely to be at one of a cluster of Sau3a restriction sites present in the cDNA clone pGR165 (section 6.2.4). As discussed earlier, <4.5kb of the mature GR1 mRNA is expected to lie upstream from the 5'end of pGR165. Preliminary Southern blot analysis of N28, using radiolabelled RNA synthesised in vitro as a probe, indicates that restriction fragments are present in N28 which do not correspond to transcription units in PCC3 cells (Reith and Stott, unpublished data). Whilst suggestive, further analysis of this clone by means of SI mapping, primer extension, DNA sequencing, and transient expression assays are necessary to determine whether this clone contains regulatory elements of the GR1 gene.

The nature of the locus (or loci) represented by clones N20.1 and N20.3 is unclear. Cross hybridisation studies will determine whether these represent the same locus. It is not known if these clones contain regions of homology to other fragments of the pGR165 cDNA clone.

Thus, along with CAD18/20 and CAD 22, genomic clones have been isolated which represent 3-5 loci with some degree of homology to the cDNA clone pGR165, in addition to the GR1 locus itself. It would be of interest to determine if transcripts from these loci could account for the additional pGR165 homologous transcripts seen in Northern blot analyses.

Section 4.3.4. 1.6U genomic clones.

Only a limited amount of data are available concerning the nature of the 1.6U locus (or loci). Southern blot analysis established the presence of four 1.6U homologous PstI-PstI restriction fragments in the PCC3 genome (fig. 3.2). It is clear that one of these genomic fragments (2.4kb) is found in clone N13.2, whilst another (760bp) is present in clones N12.1 and N13.1. The 1.6U homologous region appears to flank the vector arms in clones N1.1 and N14.1 or contains an internal SalI restriction site within the region of homology to the probe.

Restriction mapping and cross-hybridisation studies have established that N12.1 and N14.1 are overlapping clones of one locus, whilst N1.1 and N14.1 appear to represent overlapping clones of a distinct 1.6U homologous locus (L. Lazarus, unpublished data). It is not known whether the clone N13.2 overlaps with either of these loci. Clearly, more work is necessary to determine which of the 2 (or possibly 3) 1.6U homologous loci is represented by the 1.6U cDNA clone.

Chapter 5. RNA *in situ* hybridisation and fusion protein analyses of GR1.

Section 5.1. Introduction.

Section 5.1.1. Analysis of GR1 expression during mouse development.

The Northern blot analyses presented in chapter 3 indicated that GR1 expression was subject to temporal regulation during mouse development. This, along with the suggested cell type preferences of GR1 expression (section 3.2), made the expression pattern of this gene during mouse embryogenesis worthy of further study.

Even with the use of careful dissection techniques it is not possible to determine the in vivo expression pattern of a gene at the cellular level by Northern blot analyses. However, it is possible to use radiolabelled nucleic acid probes for hybridisation studies to fixed and permeabilised sections of embryonic tissues. By carrying out such RNA in situ hybridisation experiments, using pGR165-derived probes, on embryos at various stages of development, it is possible to determine the distribution of GR1-expressing cells within the embryo at all stages of development. From the known properties of the cell types expressing GR1 it might be possible to predict potential functions for this novel gene.

Whilst RNA in situ hybridisation is a powerful technique, it will give no information concerning the localisation of any putative GR1 polypeptides within cells or tissues in which the gene is active. However, this could be established by

immunocytochemical staining techniques if antibodies directed against GR1-encoded antigens were available.

Section 5.1.2. RNA *in situ* hybridisation analyses.

As discussed earlier, Northern blot analysis of gene expression is limited by the efficiency of cell separation methods. Whilst in some cases, such as testis, purification of the different cell types of a tissue is possible (Wolgemuth *et al.*, 1984), separation is generally carried out simply by physical dissection of tissues, each tissue being composed of many cell types. In addition, the small size of early mammalian embryos can present a formidable barrier in the isolation of sufficient RNA for Northern blot analyses.

RNA *in situ* hybridisation to whole embryo and tissue sections facilitates determination of the spatial and temporal expression patterns of specific mRNA sequences during development at a much higher resolution than available by Northern blot analysis, and allows easier access to early stages of embryogenesis. The usefulness of this technique depends upon the sensitivity and specificity of the hybridisation signal obtained, and it was these factors which initially limited the application of this technique to the study of repetitive elements, such as satellite DNA and ribosomal RNA (Gall and Pardue, 1971).

Early attempts to analyse expression of low copy genes made

use of tritiated DNA probes for hybridisation to purified cell populations or established *in vitro* cell lines (Harrison *et al.*, 1973, 1974; John *et al.*, 1977). Harding *et al.* (1977) used tritiated cDNA probes, prepared from total pancreas cell RNA, to study the developmental expression of pancreatic genes. However, such analyses were severely limited by the low hybridisation efficiency, estimated to be in the region of 5% (Harrison *et al.*, 1974). The subsequent efforts of a number of groups led to the development of optimum conditions for RNA *in situ* hybridisation experiments.

Treatment of cells with proteinase K, along with the use of excess probe, was found to improve hybridisation efficiency (Brahic and Haase, 1978). Sensitivity was further improved by the use of single stranded radiolabelled nucleic acid probes (Diaz *et al.*, 1981) which have been shown to give an 8-fold increase in signal strength relative to double stranded nucleic acid probes (Cox *et al.*, 1984). Radiolabelled single stranded RNA probes can be easily generated *in vitro* by use of bacteriophage RNA polymerases on a DNA template derived from plasmids containing the corresponding promoter elements (Melton *et al.*, 1984). Hydrolysis of the nucleic acid probe to an average size of 100-250 nucleotides was observed to further optimise signal strength (Hafen *et al.*, 1983; Cox *et al.*, 1984). Successful results have also been obtained using ^{35}S radiolabelled oligonucleotide probes (Meijer *et al.*, 1987).

Biotinylated nucleic acid probes have also been used in RNA *in*

situ hybridisation experiments (Singer and Ward, 1982; Brigati *et al.*, 1983). However, the sensitivity of these probes is generally not as good as that obtained on autoradiography of sections hybridised with ^{3}H or ^{35}S radiolabelled probes.

Efficient retention of RNA in sectioned tissue can be achieved by fixation and post-fixation in paraformaldehyde (Angerer and Angerer, 1981) or glutaraldehyde (Cox *et al.*, 1984). Treatment of sections with acetic anhydride was observed to improve signal:noise ratios (Angerer and Angerer, 1981), presumably by acetylation of positively charged groups which may non-specifically bind nucleic acid probe molecules. Additional improvements to the signal:noise ratio were achieved by the use of post-hybridisation RNase washes (Cox *et al.*, 1984).

The application of all these modifications led to the development of highly sensitive and reproducible techniques for RNA *in situ* hybridisation to *Drosophila melanogaster* (Hafen *et al.*, 1983) and sea urchin (Cox *et al.*, 1984) embryos. With minor modifications, (Hogan *et al.*, 1986; Wilkinson *et al.*, 1987a) RNA *in situ* hybridisation can now be reliably applied to the study of mouse embryogenesis.

Following fixation, mouse embryos can be embedded in plastic and sectioned with a cryostat (Hogan *et al.*, 1986) or dehydrated, embedded in paraffin wax and sectioned with a microtome (Wilkinson *et al.*, 1987a). The latter method results in better preservation of cell and tissue morphology. Control experiments are essential to ensure that the autoradiographic

signal seen is not due to non-specific hybridisation. Such negative controls can be provided by hybridisation of serial sections of the same tissue with either an anti-sense (complementary to the RNA species of interest) or sense strand (identical to that of RNA species of interest) radiolabelled probe, derived from the same double-stranded DNA molecule.

In recent years, the application of this technique has generated a large volume of data concerning spatial and temporal patterns of gene expression during embryogenesis. In embryos of Drosophila melanogaster, RNA in situ hybridisation has revealed that homeobox-containing genes are expressed in a segmentally restricted manner prior to morphological differentiation of segmented structures (Hafen et al., 1984a,b; Ingham et al., 1985; Bopp et al., 1986).

In the mouse, homeobox-containing genes have also been found to exhibit spatially restricted patterns of expression along an anterior-posterior axis of the embryo. Expression is generally restricted to the hind brain and thoracic regions of the central nervous system and thoracic somite-derived tissues, such as vertebrae and ribs, at specific stages of development (Awgulewitsch et al., 1986; Dony and Gruss, 1987a; Krumlauf et al., 1987; Toth et al., 1987; Utset et al., 1987; Le Mouellic et al., 1988). Localisation of Hox-1.5 expression within the three germ layers has been observed as early as 7.5 days gestation (Gaunt et al., 1986; Gaunt, 1987).

Expression of the murine retroviral-like element Etn in 3.5 days to 7.5 days gestation mouse embryos has been investigated

by RNA *in situ* hybridisation (Brulet *et al.*, 1985). Expression of the proto-oncogene c-*fos* has been found to be spatially restricted to perichondral growth regions of the cartilaginous skeleton and mesodermal web tissue of 17 days gestation mouse embryos (Dony and Gruss, 1987b). The proto-oncogene *int*-1 has been shown to be expressed within specific regions of the neural plate, and subsequently derived tissues, between 9-14.5 days of mouse embryogenesis (Wilkinson *et al.*, 1987b).
In order to investigate further the expression pattern of GR1 during mouse development, the cDNA clone pGR165 was used to generate ^{35}S radiolabelled RNA probes specific for the GR1 mRNA. These probes were used in a series of RNA *in situ* hybridisation experiments to sections of mouse embryos from 7.5 days to 17.5 days gestation. GR1 expression in adult testis was also analysed by this method.

Section 5.1.3. Raising polyclonal antisera against putative GR1 proteins.

At the initiation of antibody studies, the only representation of the GR1 gene product available for experimental analysis was the potential amino acid sequence encoded within the cDNA clone pGR165. Determination and analysis of the nucleotide sequence of this cDNA clone would reveal any polypeptide-encoding capacity. This information could be used to produce synthetic peptides corresponding to regions of the predicted

GR1 polypeptide which, in turn, could be used to raise polyclonal antisera (Walter, 1986).

A potentially more rapid approach to the production of GR1 antisera was offered by the expression of any polypeptide-encoding capacity of the cDNA clone pGR165 *in vivo*. Polypeptides representing this pGR165-encoded amino acid sequence could then be used directly for immunisation. This strategy could be easily carried out by the synthesis of β-galactosidase-GR1 fusion proteins in *E.coli*.

The *lac* operon of *E.coli* represents one of the best characterised gene regulatory systems of prokaryotes (see Beckwith and Zipser, 1970 for review). The first structural gene of the *lac* operon, *lacZ*, encodes the enzyme β-galactosidase. *lacZ* has been used extensively in experiments designed to express and isolate DNA restriction fragments, in *E.coli*, corresponding to antigens against which antibodies are available.

β-galactosidase activity is known to be retained when *lacZ* is fused with other genes (Guarente and Ptashne, 1981; Rose *et al.*, 1981). A number of plasmid vectors have been generated which facilitate the insertion of genomic or cDNA restriction fragments at the 5′ end of *lacZ* (Gray, *et al.*, 1982; Koenen *et al.*, 1982; Ruther *et al.*, 1982). In these vectors, expression of β-galactosidase activity occurs only when the open reading frame of the inserted DNA is in the same frame as that of the *lacZ* gene. Other vectors such as the bacteriophage lambda vector gt11 (Young and Davis, 1983) and the pUR288-292 series

of plasmid vectors (Ruther and Muller-Hill, 1983) facilitate cloning of restriction fragments at the 3′ end of lacZ. This is particularly useful in expressing cDNA clones since many cDNA molecules will contain the translational termination codon of the open reading frame of interest. Moreover, it has been observed that insertion of DNA restriction fragments into the 5′ end of lacZ can give rise to unstable fusion proteins, whereas insertion into the 3′ end often results in the synthesis of stable fusion proteins (Stanley, 1983). However, 5′ insertions are not always unstable (Gray et al., 1982).

Together, these vectors have allowed the construction of recombinant genomic and cDNA libraries, in E.coli, in which the cloned DNA molecules can be expressed as β-galactosidase fusion proteins. Screening such libraries with antibody probes using either autoradiographic (Broome and Gilbert, 1978; Erlich et al., 1978; Kemp and Cowman, 1981) or histochemical staining (Koenen et al., 1982; Ruther et al., 1982; Ruther and Muller-Hill, 1983) techniques can result in the isolation of cDNA or genomic clones corresponding to an antigen of interest, in the absence of any amino acid sequence data for that antigen.

This technology has also been utilised in experiments designed to raise antisera against antigens encoded by DNA restriction fragments. β-galactosidase fusion proteins have been used successfully to raise polyclonal antisera against the malF gene of E.coli (Shuman et al., 1980) and Herpes Simplex virus I thymidine kinase (Weinstock et al., 1983). The pUR290-292

series of plasmid vectors (Ruther and Muller-Hill, 1983) have been used to raise polyclonal and monoclonal antisera against novel epitopes on SV40 large T-antigen (Mole and Lane, 1985). This latter case illustrates the utility of the pUR series to determine which, if any, reading frame of a given strand of a cDNA molecule represents a large open reading frame. Three vectors, pUR290, pUR291 and pUR292 are designed such that each reading frame of a given strand of a DNA restriction fragment can be juxtaposed with the open reading frame of the lacZ gene.

The growth disadvantage imposed by over-production of β-galactosidase (Koenen et al., 1985), and the potential toxicity of some β-galactosidase fusion proteins, demands that the expression of fusion proteins is tightly regulated. This can be achieved in the pUR series by use of a host E.coli strain, such as BMH71-18 (section 2.2), which over-produces lac repressor (Koenen et al., 1985). Culture of such E.coli cells carrying a pUR plasmid in medium containing the lac inducer IPTG will result in induction of expression of the fusion protein.

Thus, by comparison of the size of the β-galactosidase fusion protein produced by each recombinant plasmid in IPTG-grown cultures, it is possible to determine which of the three potential reading frames of the DNA restriction fragment of interest contains a long open reading frame. The fusion protein produced by this clone can then be used as an immunogen.

Given that only one strand (the sense strand) of a cDNA molecule represents the nucleotide sequence of the original mRNA, only plasmid constructs in which this strand is in frame with lacZ need be analysed. Determination of the sense/ anti-sense orientation of the cDNA clone can be determined by Northern blot analyses using strand-specific probes of the cDNA clone.

Antiserum raised against the fusion protein representing the longest open reading frame of the sense strand of the cDNA clone pGR165 would be expected to contain activities directed against both β-galactosidase and GR1-encoded regions of the fusion protein. Any anti-GR1 activity could be detected by comparison of Western blot analysis of crude protein extracts and immunocytochemical staining analyses of GR1-expressing and non-expressing cell types. The anti-GR1 antiserum could then be used for immunocytochemical staining of embryonic tissues and cells to determine the location of the GR1 protein within the developing embryo and within GR1-expressing cells.

Data derived from both antibody studies and RNA in situ nucleic acid hybridisations, along with analysis of amino acid sequence data (chapter 6), would prove invaluable in the determination of the role(s) played by GR1 during mouse development.

Section 5.2. Results.

Section 5.2.1. Transcriptional orientation of pGR165.

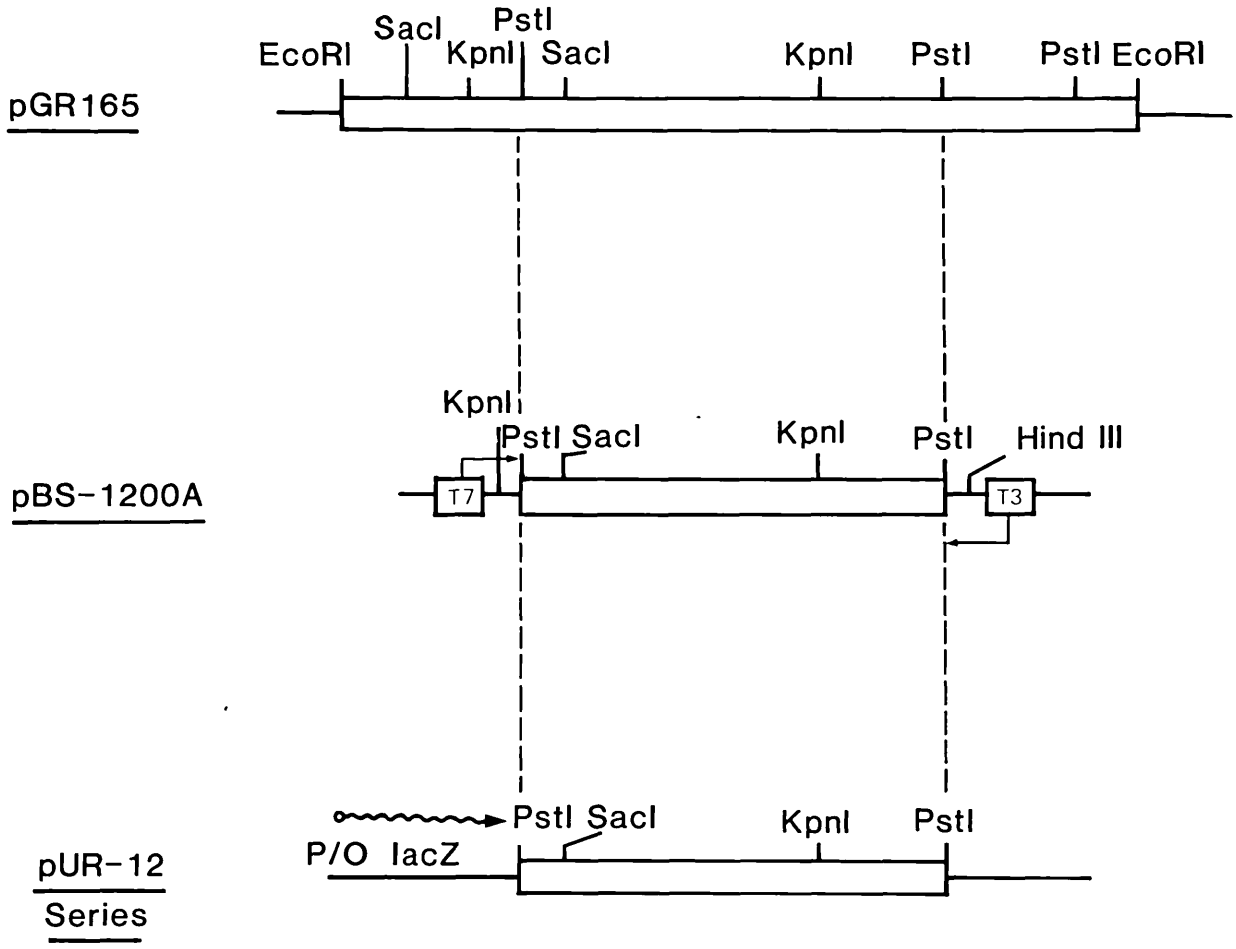
Only one strand of a double-stranded cDNA clone will represent the coding sequence of the original mRNA from which it was synthesised. Given the complementarity of the two strands of a cDNA clone, it would be expected that only one strand (anti-sense) would hybridise to the mRNA. This information was critical for RNA *in situ* hybridisation analyses of GR1 expression using single-stranded probes derived from the cDNA clone pGR165. Also, in the absence of the complete nucleotide sequence of pGR165 there was no knowledge concerning the presence of any potential polypeptide encoding regions within this 2.5kb cDNA clone. Clearly, it was desirable to analyse only those potential open reading frames found in the sense strand of pGR165.

In order to determine the sense/anti-sense orientation of pGR165, radiolabelled strand-specific RNA probes, synthesised *in vitro* from the plasmid pBS1200A (fig. 5.1), were used in Northern blot analyses of total cellular RNA isolated from PCC3 cells or their differentiated derivatives.

As shown in fig. 5.2a, only one of the strand-specific RNA probes generated from pBS1200A, that synthesised by T3 polymerase, hybridised to the characteristic GR1 mRNA species found in PCC3 cells but not differentiated cell types. A faint

Figure 5.1. Construction of lacZ-GR1 fusion protein encoding plasmids.

Restriction maps of the cDNA clone pGR165 and the plasmid subclones, pBS1200A and the pUR-12 series (290-12, 291-12 and 292-12) of lacZ-GR1 fusion proteins, are shown. T7 and T3 represent transcriptional promoter elements responsive to RNA polymerases of T7 and T3 bacteriophage respectively. P/O = promoter and operator elements of lacZ. -> = direction of transcription mediated by T3 or T7 polymerases and of the lacZ-GR1 fusion constructs.


pGR165
EcoRI
SacI
KpnI
PstI
SacI
KpnI
PstI
PstI
EcoRI
pBS-1200A
KpnI
PstI
SacI
KpnI
PstI
Hind III
T7
T3
pUR-12
Series
P/O
lacZ
PstI
SacI
KpnI
PstI

smear of hybridisation was observed in both PCC3 and differentiated PCC3 RNA samples with the sense RNA probe (fig. 5.2a).

Thus, it was concluded that the DNA sequence of the upper strand of pGR165, as shown in fig. 5.1, represents the nucleotide sequence of the GR1 mRNA. Transcription is assumed to occur in the 5′ to 3′ direction i.e. left to right as the cDNA is shown in fig. 5.1.

Section 5.2.2. Production of β-galactosidase-GR1 fusion proteins in E.coli.

Having established the transcriptional orientation of pGR165, it was now possible to determine if the sense-strand of this cDNA clone contained any large open-reading frames against which antisera could be raised. As discussed earlier (section 5.1.2), the pUR series of plasmid vectors can be used for this purpose.

Initial experiments involved subcloning the 2.5kb EcoRI-EcoRI restriction fragment of the cDNA insert in pGR165 (fig. 5.1) into each of the pUR vectors, pUR290, pUR291 and pUR292. The cohesive ends of the 2.5kb EcoRI fragment of pGR165 were converted to blunt ends (section 2.16.4) and ligated with each of the three pUR vectors linearised with SalI and blunt ended (section 2.16.4). This strategy was designed to ensure that no translational termination codons would be created, in any

frame, at the site of ligation.

Analysis of numerous recombinant clones in each of the vectors pUR290, pUR291 and pUR292 resulted in the isolation of two classes of clones. In the first class, the anti-sense strand of the 2.5kb restriction fragment of pGR165 was in the same orientation as lacZ. The second class, despite hybridising to radiolabelled 2.5kb EcoRI fragment of pGR165, gave restriction maps quite dissimilar to those predicted for clones containing this fragment in either orientation (data not shown). No clones containing the sense strand of the restriction fragment of pGR165 in the same orientation as lacZ were isolated. SDS-PAGE analysis showed cultures of clones of the first class to contain an IPTG inducible protein the same size as β-galactosidase (data not shown). No IPTG-inducible protein bands were seen in cultures of the second class of clones.

Given the lack of success in subcloning the complete cDNA insert of pGR165, a new strategy was adopted, using the 1.1kb PstI-PstI restriction fragment of pGR165, in attempts to generate β-galactosidase-GR1 fusion proteins (fig. 5.1).

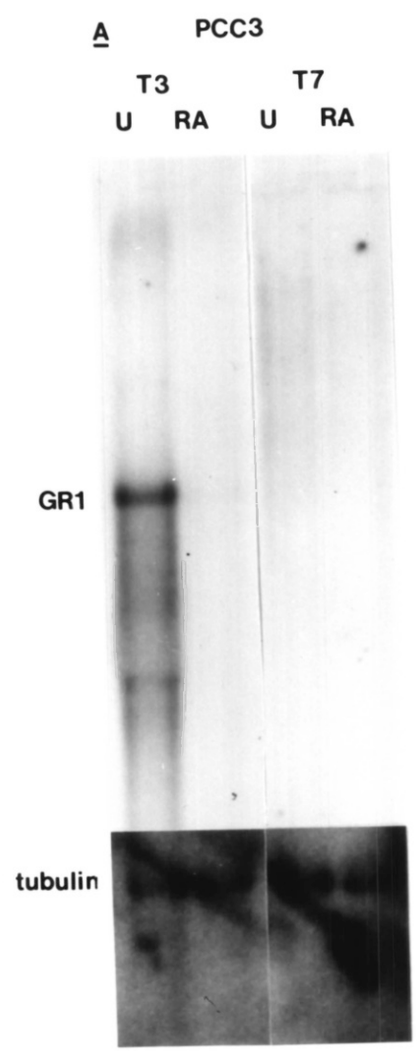
The PstI restriction enzyme site in the pUR vectors 290, 291 and 292 facilitated the cloning of this fragment at the 3′ end of lacZ. Additionally, the 1.1kb PstI-PstI restriction fragment did not contain the moderately repetitive element present in pGR165 (section 4.2.2). This element may have been responsible for the apparent instability of the complete cDNA molecule in the sense orientation in the pUR vectors.

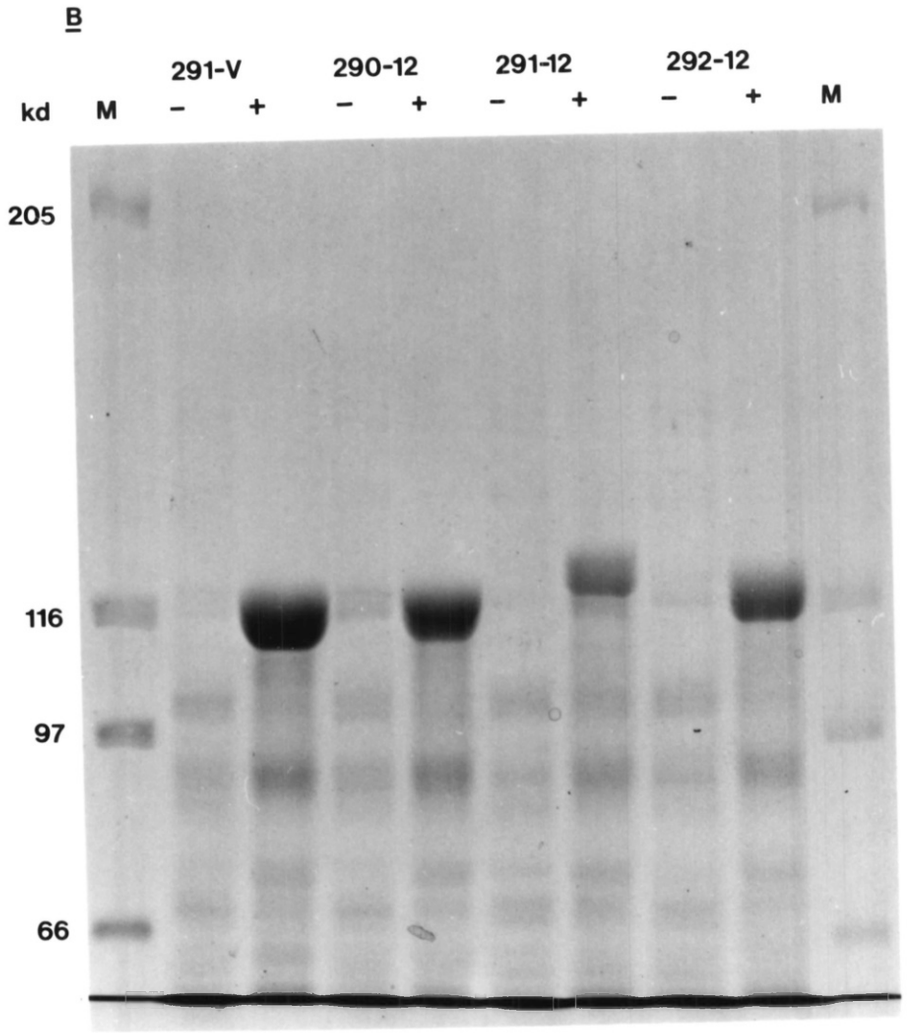
Figure 5.2. Transcriptional orientation and fusion protein analysis of pGR165.

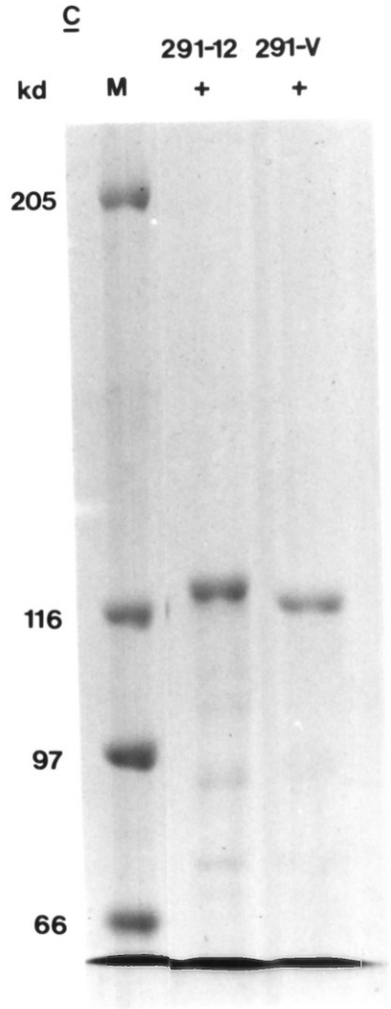
A. Northern blot analysis. 10ug of total cellular RNA isolated from PCC3 cells (U) or PCC3 cells treated for 10 days with retinoic acid (RA) was electrophoresed in a 1%(w/v) agarose-formaldehyde gel (section 2.18) and transferred to nitrocellulose. Filters were hybridised with radiolabelled RNA probes synthesised *in vitro* from pBS1200A linearised with *Eco*RI using T3 polymerase (specific activity $1.8x10^8$cpm/ug; $1.6x10^6$cpm/ml used) (T3), or pBS1200A linearised with *Hin*dIII using T7 polymerase (specific activity $2.2x10^8$cpm/ug; $2x10^6$cpm/ml used) (T7). Both filters were washed to 0.1xSSC, 0.1%(w/v) SDS at 65^oC and exposed for 2 days. The filters were subsequently hybridised with oligolabelled *Pst*I fragment of pBSBT (spec. activity $1.2x10^8$cpm/ug; 10^6cpm/ml used), washed to 0.1xSSC, 0.1%(w/v) SDS at 65^oC and exposed for 16 hours.

B. 10ul samples of crude protein extracts (section 2.12) of *E.coli* BMH71-18 cultures bearing various plasmid constructs (section 5.2.2) were subjected to 5% SDS-PAGE and stained with Coomassie blue (section 2.15.4). - and + = cultures grown in the absence or presence of IPTG. 291-V= pUR291 vector plasmid. 290-12, 291-12, 292-12 = each of the pUR-12 series of plasmids (fig. 5.1). kd= kilodaltons. M= marker proteins: rabbit myosin (205kd), *E.coli* β-galactosidase (116kd), rabbit phosphorylase b (97kd), and bovine albumin (66kd).

C. As for b. but showing Coomassie stained 5% SDS-PAGE of 1ul samples of crude protein extract.


A
PCC3
T3
T7
U
RA
U
RA
GR1
tubulin


B
291-V
290-12
291-12
292-12
kd
M
–
+
–
+
–
+
–
+
M
205
116
97
66


C
291-12
291-V
kd
M
+
+
205
116
97
66

Moreover, the element may encode an antigenic site common to a number of polypeptides otherwise unrelated to GR1.

The 1.1kb PstI-PstI restriction fragment of pGR165 was subcloned into the PstI site of pUR290, 291 and 292 and the orientation of insert fragments in recombinants determined by restriction map analysis (data not shown). Three clones, designated 290-12, 291-12 and 292-12, were found to contain single copies of the pGR165 restriction fragment in which the GR1 sense strand was in the same transcriptional orientation as the lacZ gene (fig. 5.1 and data not shown).

An essential feature of lacZ fusion vectors is that the production of fusion proteins is normally repressed in the host E.coli strain bearing the recombinant plasmid (section 5.1.2). Transcription of the lacZ gene in the pUR series of plasmid vectors, propagated in E.coli BMH71-18, is normally repressed by over-production of lac repressor. It is possible to induce synthesis of the galactosidase fusion protein by growing cultures in the presence of the lac inducer, IPTG.

Samples of crude protein extract (section 2.12), prepared from uninduced and induced cultures of E.coli bearing each of the plasmids 290-12, 291-12, 292-12 and the vector pUR291, were subjected to SDS-PAGE and proteins detected by staining with Coomassie blue (section 2.15.4). As shown in fig. 5.2b, a single abundant band was seen to be induced in IPTG-treated cultures of E.coli bearing each of the four plasmids. Cultures of the host E.coli strain carrying no pUR plasmid did not show induction of a similar sized protein band (data not shown).

From the rate of migration of the IPTG-inducible band, relative to that of purified β-galactosidase (116kd), it was apparent that only one of the cultures, that carrying 291-12, contained an IPTG-inducible band detectably larger than that produced in E.coli cultures carrying the plasmid vector pUR291 (fig. 5.2b). This band was estimated to be approximately 121kd in size (fig.5.2c and data not shown). This fusion protein was detected in cultures of three independent clones of the sense strand of pGR165 in pUR291 (data not shown). Whilst stable in cultures grown at 27°C, considerable degradation was observed in cultures grown at 37°C (data not shown).

Thus, the data suggested that only one of the three reading frames of the sense strand of the 1.1kb PstI-PstI fragment of pGR165 contained a large open reading frame. On this basis, polyclonal rabbit antisera were raised against the 291-12 fusion protein.

Section 5.2.3. Polyclonal rabbit antisera against the 291-12 fusion protein.

The abundant 121kd IPTG-inducible protein synthesised in E.coli BMH71-18 carrying the recombinant plasmid 291-12 (section 5.2.2) was used as an immunogen for two rabbits as described in section 2.24. (Injections kindly performed by N. Clark, National Institute for Medical Research).

Production of antibodies against the 291-12 fusion protein was

monitored by Western blot analysis (section 2.25) of crude protein extracts of E.coli cultures bearing plasmid 291-12 grown in the presence of IPTG. Only that area of the filter corresponding to the IPTG-inducible protein band used as immunogen was used for Western blot analyses.

Whilst serum from the pre-immune animals did not contain any detectable activity against the 291-12 protein, that from each of the immunised animals reacted against this fusion protein. In rabbit 1, anti-fusion activity was detected at 3 days, and peaked 12 days after the final injections. At this time a 500-fold dilution of antiserum gave a clear reaction to the 291-12 fusion protein (data not shown). The other animal also showed a response to the immunogen, but in this case the titre was considerably lower (data not shown). Subsequent analysis was carried out using the highest titre antiserum of the first animal.

Thus far, only a reaction against the entire immunogen had been established. Whilst indicating that the immunisation procedure had been successful, it was not clear what, if any, antibody activity had been raised against the GR1-encoded region of the fusion protein. Western blot analysis of uninduced and IPTG-induced cultures of E.coli carrying plasmid 291-12 confirmed the presence of antisera directed against the major IPTG-inducible protein band in these cultures (data not shown). However, Western blot analyses of crude protein extracts of murine cell lines PCC3 and SV3T3 C138, along with PCC3 cells treated with retinoic acid for 10

days (section 2.3), using the 291-12 antiserum, failed to detect any PCC3-specific protein band, even when a 100-fold dilution of 291-12 antiserum was used (data not shown). Moreover, immunocytochemical staining of PCC3 and SV3T3 C138 cells with the 291-12 antiserum failed to reveal any activity specifically against antigens present in PCC3 cells (data not shown).

Thus, whilst antiserum with activity against the fusion protein 291-12 had been raised, there was no evidence to suggest the presence of any antibodies directed against GR1-encoded antigens, either in PCC3 cell extracts or intact PCC3 cells.

Section 5.2.4. RNA *in situ* hybridisation analyses.

As mentioned earlier (section 5.1.1), RNA *in situ* hybridisation represents a very powerful technique for the determination, at the cellular level, of the expression pattern of a given gene during embryogenesis. Northern blot analyses had established that GR1 was expressed throughout the period of 9.5 days to 16.5 days gestation (section 3.2.9).

Thus, in parallel with attempts to raise polyclonal antisera against GR1-encoded antigens (section 5.2.3), a series of RNA *in situ* hybridisation experiments were carried out to determine the distribution of GR1 mRNA in normal mouse embryos between 7.5 days and 17.5 days of gestation. In addition to

analysing GR1 expression at a higher degree of resolution than possible by Northern blot analysis, such studies would reveal the pattern of GR1 expression at earlier stages of embryogenesis than previously investigated.

Embryos were dissected from the uterus and fixed, permeabilised, embedded, and sectioned as described in section 2.19. Sections were hybridised with ^{35}S radiolabelled strand-specific RNA probes representing each strand of the pGR165 subclone pBS1200A (fig. 5.1). Following hybridisation, the sections were washed (section 2.20.5.) and localisation of radiolabelled probe determined by autoradiography (section 2.21.3).

In all cases, only the probe generated by T3 polymerase (fig. 5.1), gave a reproducible hybridisation signal. No hybridisation signal was detected with RNA probes generated from pBS1200A using T7 polymerase (see below). Given the known sense/anti-sense orientation of pBS1200A (section 5.2.1), it was concluded that the hybridisation signal seen with RNA probes generated from this plasmid using T3 polymerase was due to the presence of GR1 mRNA. In all experiments, probes representing the sense-strand of the GR1 transcription unit were used as a control for non-specific hybridisation.

Some batches of [$\alpha^{35}S$] UTP were found to specifically inhibit the activity of T3 polymerase (data not shown). Thus, in some of the experiments, GR1 anti-sense RNA probes were generated from a template of plasmid pBS1200B using T7 polymerase. Like

pBS1200A (fig. 5.1), pBS1200B contains a single copy of the 1.1kb PstI-PstI restriction fragment of pGR165 in the vector pBSM13^{+}. However, the orientation of the PstI fragment in pBS1200B is the opposite to that found in pBS1200A (data not shown).

These experiments revealed that GR1 is expressed at a number of discrete sites in the developing embryo. Both the stages of development and cell types within embryonic tissues were determined by comparison of embryo sections with the data of Rugh (1968) and Theiler (1972).

5.2.4.(i) GR1 expression in extraembryonic tissues.

High levels of GR1 RNA were detected in a variety of extraembryonic tissues at all stages of development investigated (fig. 5.3; fig. 5.4; data not shown). Within the 7.5 days gestation conceptus, GR1 expression was detected in trophectoderm-derived giant cells and diploid trophoblast of the ectoplacental cone as well as in chorion, amnion, visceral endoderm and parietal endoderm (fig. 5.3; data not shown). At 8.5 days gestation, an additional site of expression was observed in the mesoderm-derived allantois (fig. 5.5a).

At later stages of development, these extraembryonic tissues contribute to the placenta and membranes around the fetus (Theiler, 1972). As seen at 14.5 days gestation (fig. 5.4), high levels of GR1 RNA were detected in visceral yolk sac and

the trophoblastic labyrinth layer of the placenta. Within the outer spongiotrophoblast, only the trophectoderm giant cells were seen to express GR1. The remainder of the spongiotrophoblast, of maternal origin (Rossant and Croy, 1985), and the uterine muscle layer, did not express this gene (fig. 5.4c).

5.2.4.(ii) GR1 expression during gastrulation.

Between 7.5 days and 10.5 days gestation, GR1 expression within the mouse embryo was detected in embryonic mesoderm and at restricted sites in ectodermal tissues. At 7.5 days gestation (fig. 5.3), little, or no, GR1 RNA was detectable in embryonic ectoderm. In contrast, GR1 expression was apparent in the mesodermal cells arising as the primitive streak between embryonic ectoderm and the overlying visceral endoderm (fig. 5.3).

Such an expression pattern was also seen at 8.5 days gestation (fig. 5.5). In longitudinal section (fig. 5.5a), the primitive streak at the posterior end of the embryo, and early somites, were seen to express GR1 (fig. 5.5a). At the anterior end, GR1 RNA was detected within the head mesenchyme and in other mesenchymal cells underlying the developing neural folds (fig. 5.5a,b,c). GR1 expression was seen in paraxial mesoderm cells of more caudal regions of the neural tube (fig. 5.5d,e) as well as in specific areas of neural ectoderm, notably at the

Figure 5.3. GR1 expression in 7.5 days gestation mouse conceptus.

Longitudinal sections of 7.5 days gestation mouse conceptuses were prepared (section 2.19) and hybridised with sense (A,C) and anti-sense (B,D) ^{35}S radiolabelled RNA probes prepared from pBS1200A or pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours.

A and B: 100x magnification of serial sections of the same conceptus.

C and D: 250x magnification of serial sections of a second conceptus.

A= amnion, C= chorion, E= embryonic ectoderm, Epc= ectoplacental cone, PS= primitive streak, VE= visceral endoderm.

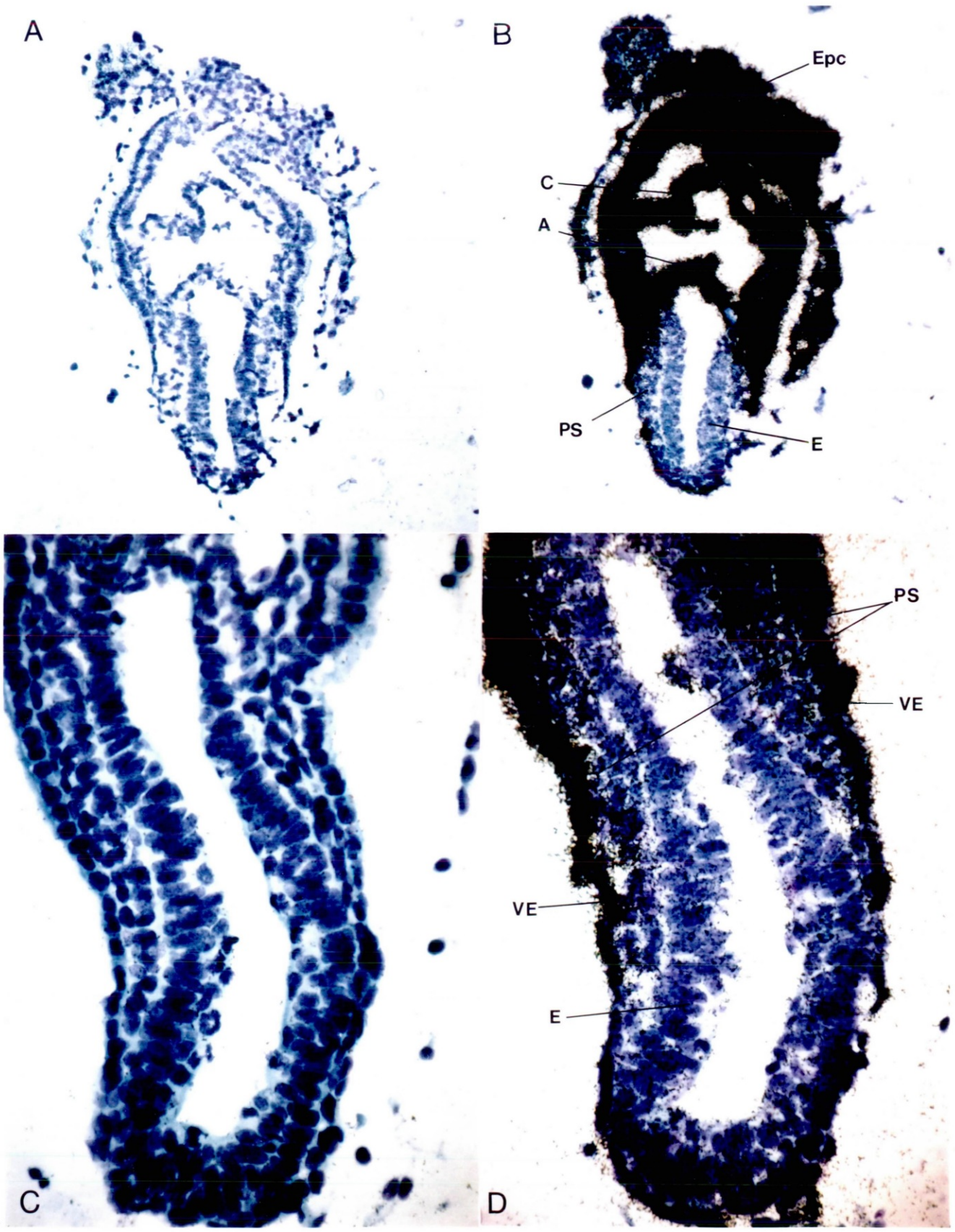

A
B
Epc
C
A
PS
E
C
PS
VE
VE
E
D

Figure 5.4. GR1 expression in mid-gestation placenta.

Transverse sections of 14.5 days gestation conceptuses were prepared (section 2.19) and hybridised with sense (B) and anti-sense (A,C) ^{35}S radiolabelled RNA probes prepared from pBS1200A or pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours.

A: 40x magnification.

B and C: 250x magnification of serial sections to that in A.

GC= trophoblast giant cells, L= Labyrinth, ML= muscle layer, Sp= spongiotrophoblast, VS= venous sinus, VYS= visceral yolk sac.

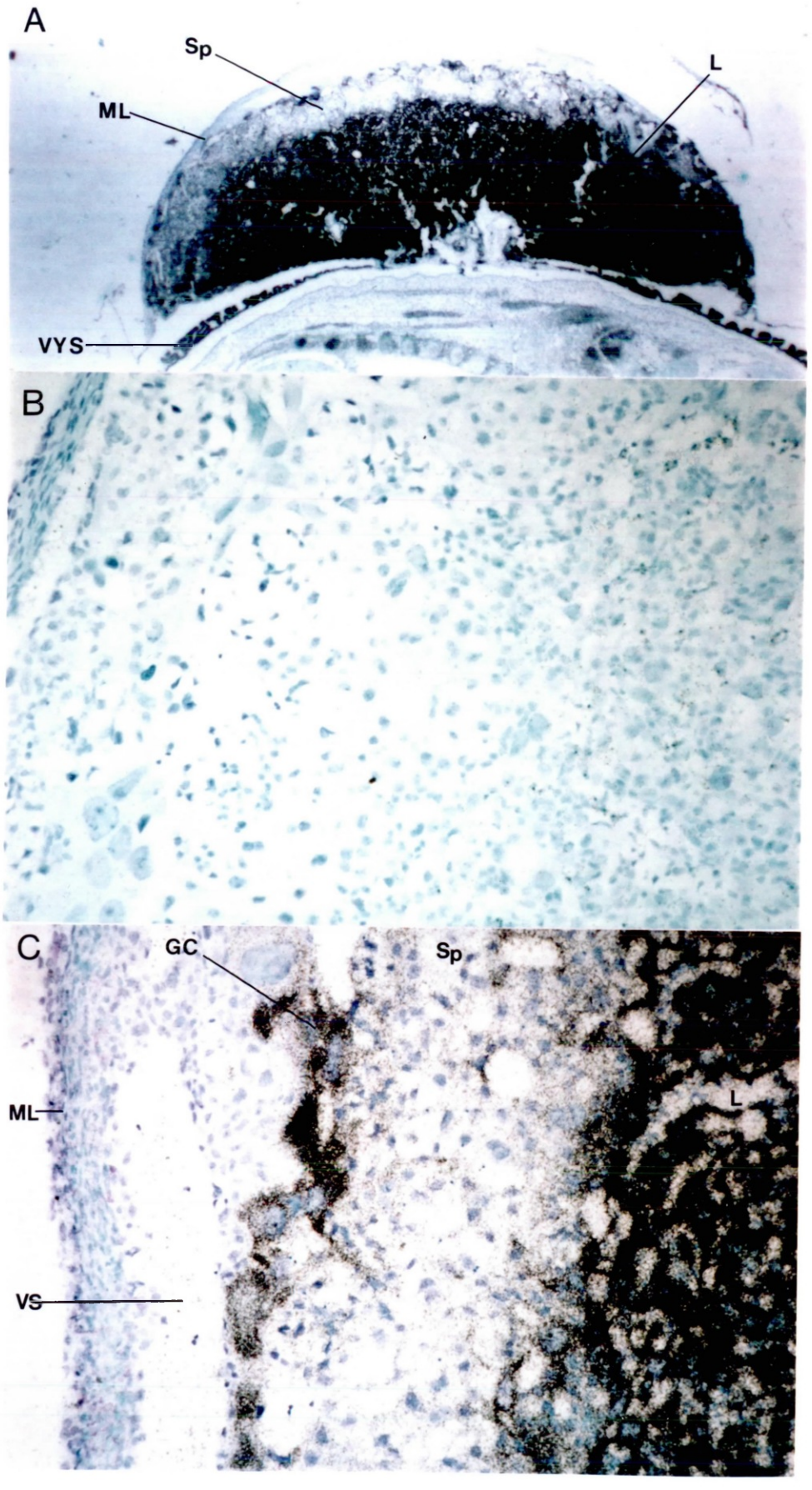

A
Sp
L
ML
VYS
B
C
GC
Sp
ML
L
VS

Figure 5.5. Localisation of GR1 RNA in 8.5 days gestation mouse embryo.

Longitudinal (A) and transverse (B,C,D,E) sections of 8.5 days gestation mouse conceptuses were prepared (section 2.19) and hybridised with sense (D) and anti-sense (A,B,C,E) ^{35}S radiolabelled RNA probes prepared from pBS1200A or pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours.

In A, the embryonic tissue has separated from the overlying visceral endoderm near the anterior end of the primitive streak. B and C represent transverse sections through the developing head folds, B being more cranial than C. D and E represent serial transverse sections of neural folds caudal to that in C.

All sections at 100x magnification.

A= amnion, Al= allantois, Epc= ectoplacental cone, F= foregut, GC= trophoblast giant cells, M= mesenchyme, N= neural folds, PS= primitive streak, S= somites, VE= visceral endoderm.

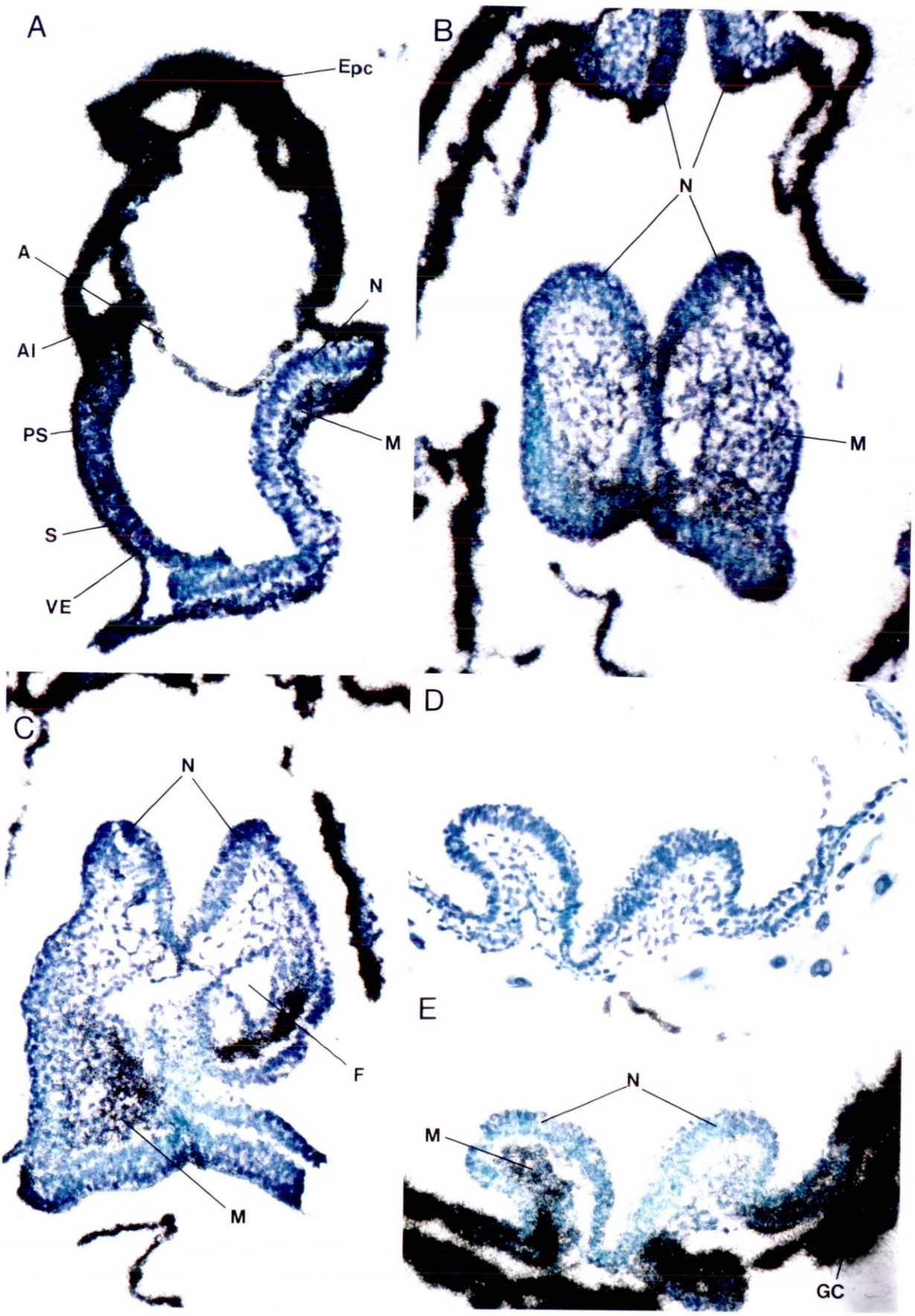

A
Epc
A
N
AI
PS
M
S
VE
B
N
M
C
N
F
M
D
E
N
M
GC

Figure 5.6. GR1 expression in 9.5 days gestation mouse embryo.

Transverse sections of 9.5 days gestation mouse conceptuses were prepared (section 2.19) and hybridised with sense (B) and anti-sense (A,C) ^{35}S radiolabelled RNA probes prepared from pBS1200A or pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours. A represents the head, and C the trunk, regions of an embryo seen in a single transverse section. B represents a serial section of the trunk region of the same embryo.

All sections are at 100x magnification.

Ca= caudal artery, Cm= coelom, Cv= caudal vein, H= hindgut, M= mesenchyme, Mec= mesencephalon, Myc= myelencephalon, Nt= neural tube.


A
M
Myc
Mec
B
C
Nt
Ca
H
M
Cm
Cv

tips of the developing neural tube (fig. 5.5b,c).

Expression in the head region was also seen at 9.5 days gestation when both head mesenchyme and specific areas within the mesencephalon and myelencephalon were found to contain GR1 RNA (fig. 5.6a). In the trunk region at 9.5 days gestation, GR1 expression was seen in the mesenchymal cells surrounding the coelom, and lower levels of expression were apparent in some areas of the neural tube (fig. 5.6c).

5.2.4.(iii) GR1 expression in sclerotome and derived structures.

The expression of GR1 within the presomitic mesenchyme of the primitive streak at 7.5 days and 8.5 days gestation was also seen at 10.5 days gestation, where the tail bud mesenchyme was found to express GR1, whilst the limiting ectoderm was essentially negative for GR1 expression (fig. 5.8a). However, following condensation and differentiation of presomitic mesenchyme to form a mature somite, GR1 expression was found to localise exclusively to the sclerotome cells, the dermamyotome being negative (fig. 5.8b).

The sclerotome of the somites subsequently gives rise to vertebrae. As shown in figure 5.7, at 12.5 days gestation the sclerotome of each somite has segregated to form densely-packed posterior, and loosely-packed anterior, sclerotome halves (Gruneberg, 1963; Theiler, 1972; fig.5.7b).

Figure 5.7. Origin of vertebral segments.

The anterior-posterior segmentation within the sclerotome, and resegmentation involved in the generation of vertebrae, is illustrated (after Gruneberg, 1963). A vertebral segment is formed from the posterior half of one sclerotome and the anterior half of the adjacent sclerotome.

A approximates to 10.5 days gestation, B to 12.5 days gestation and C to 17.5 days gestation in the mouse (Rugh, 1968).

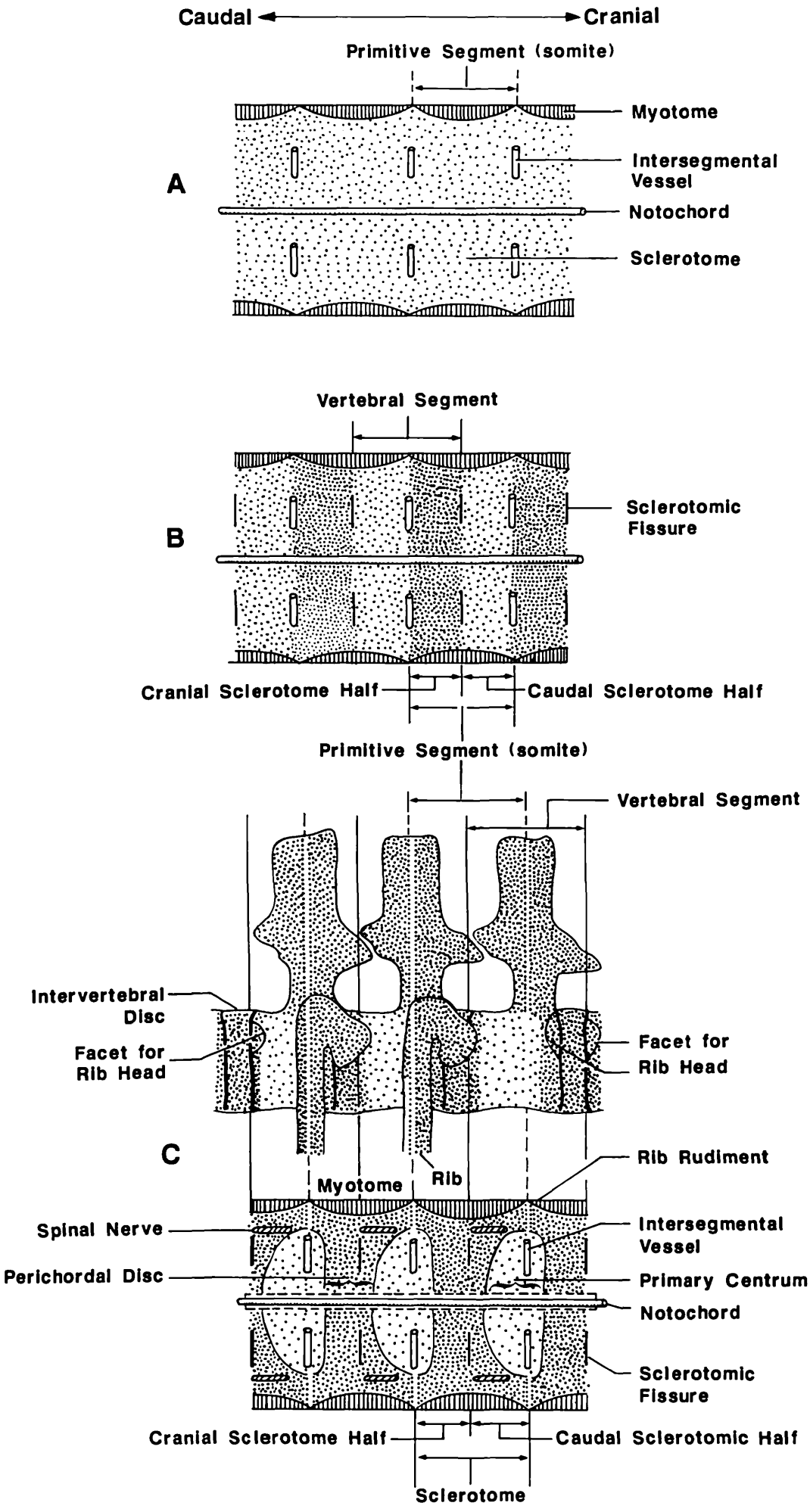

Caudal
Cranial
Primitive Segment (somite)
A
Myotome
Intersegmental Vessel
Notochord
Sclerotome
Vertebral Segment
B
Sclerotomic Fissure
Cranial Sclerotome Half
Caudal Sclerotome Half
Primitive Segment (somite)
Vertebral Segment
Intervertebral Disc
Facet for Rib Head
Facet for Rib Head
C
Rib Rudiment
Myotome
Rib
Spinal Nerve
Intersegmental Vessel
Perichordal Disc
Primary Centrum
Notochord
Sclerotomic Fissure
Cranial Sclerotome Half
Caudal Sclerotomic Half
Sclerotome

Figure 5.8. Localisation of GR1 expression in sclerotome and derivatives.

Longitudinal (A,B,F,G) and transverse (C,D,E) sections of 10.5 days (A,B), 12.5 days (C,D,E,F,) and 14.5 days (G) gestation mouse embryos were prepared (section 2.19) and hybridised with sense (C) and anti-sense (A,B,D,E,F,G) ^{35}S radiolabelled RNA probes prepared from pBS1200A or pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours.

A: 40x magnification.

B: same section as A but 100x magnification.

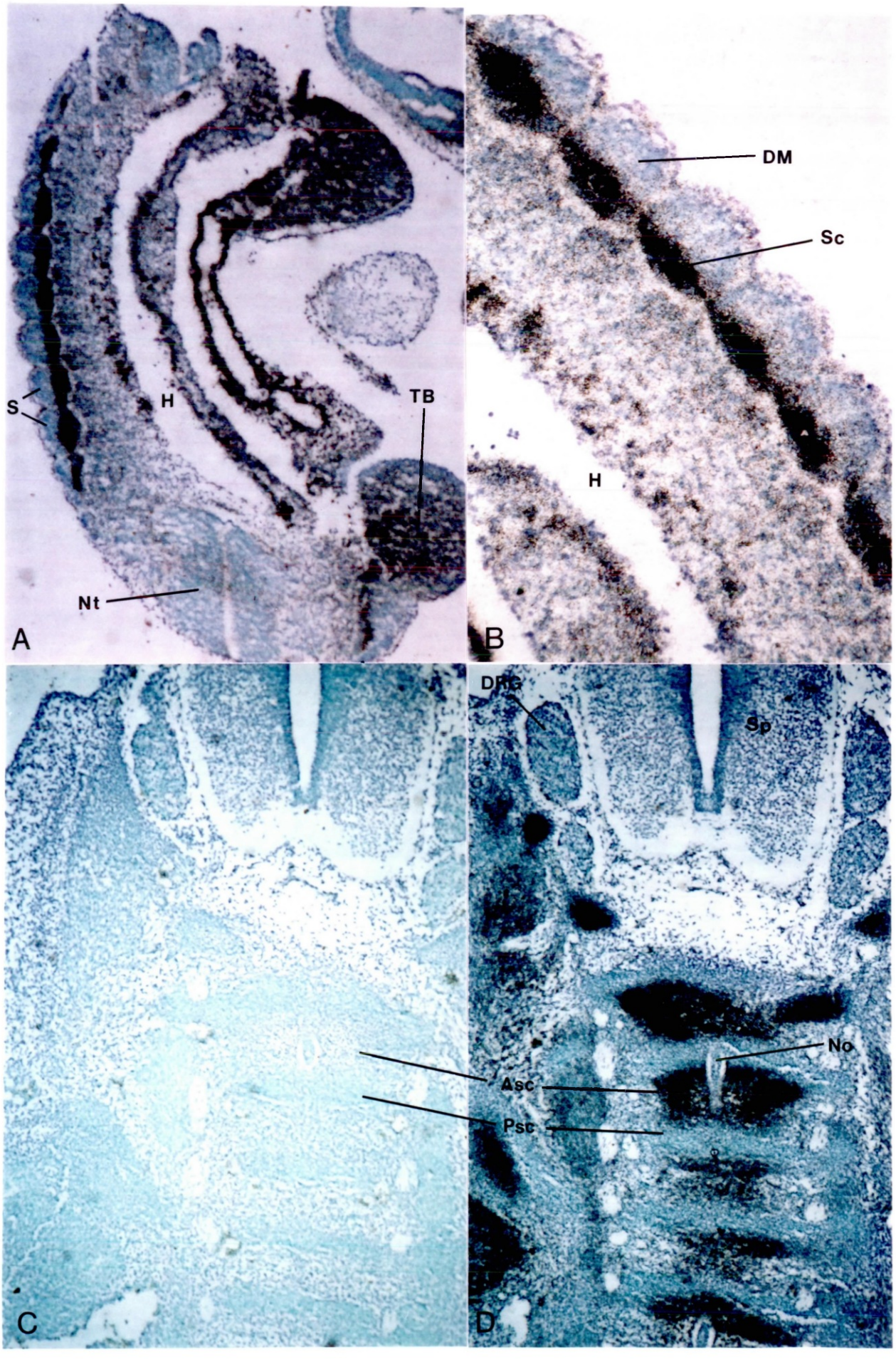
C,D,E,F,G: 100x magnification.

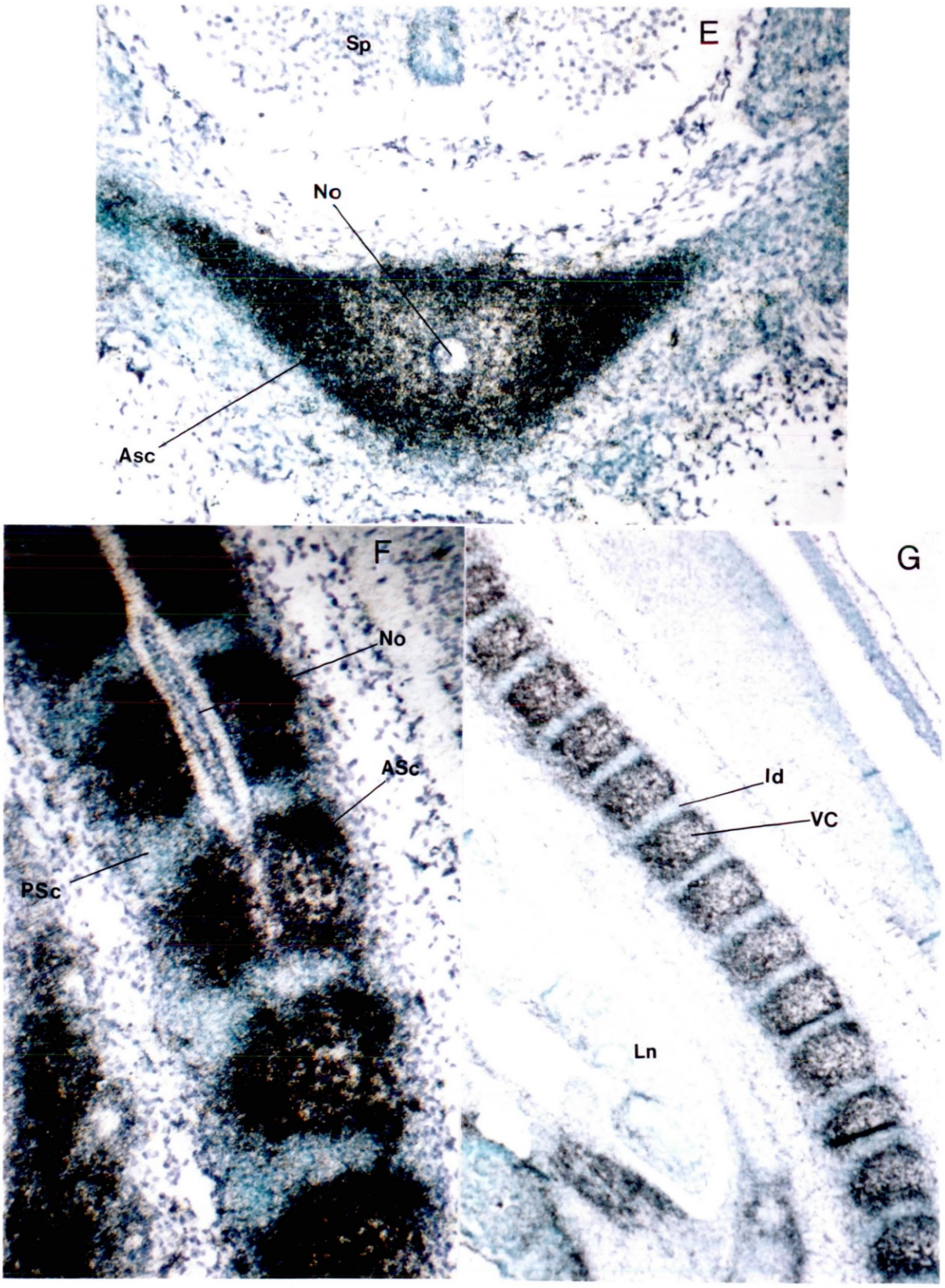
C and D are serial sections from the same embryo.

Asc= anterior sclerotome, DM= dermamyotome, DRG= dorsal root ganglion, H= hindgut, Id= intervertebral disc, Ln= lung, No= notochord, Nt= neural tube, Psc= posterior sclerotome, S= somite, Sc= sclerotome, Sp= spinal cord, TB= tail bud, VC= vertebral cartilage.

A,B and C on facing page.

D,E and F, overleaf.


S
H
TB
Nt
A
DM
Sc
H
B
DRG
Sp
No
Asc
Psc
C
D


E
Sp
No
Asc
F
No
ASc
PSc
G
Id
VC
Ln

Subsequently, a vertebral segment is formed by means of a re-segmentation of the sclerotome mesenchyme across the inter-segmental border (fig. 5.7c).

Such an arrangement ensures that the adjacent axial muscle spans two vertebrae, facilitating bending movements of the vertebral column. In this model, the anterior sclerotome half forms the vertebral body, whilst the posterior sclerotome half of the adjacent somite forms the intervertebral disc (fig. 5.7).

As seen in fig. 5.8 c,d,e and f, at 12.5 days gestation GR1 expression localises to the loosely packed anterior sclerotome half, the posterior sclerotome half being negative. At 14.5 days gestation, the anterior sclerotome half is chondrified and developing into the vertebral cartilage. It is seen that GR1 expression is retained in this tissue whilst the posterior sclerotome-derived intervertebral disc cells remain negative for GR1 expression (fig. 5.8f). It should also be noted that the notochord was found not to express GR1.

5.2.4.(iv) GR1 expression in other embryonic skeletal tissues.

In addition to being expressed within the developing vertebral cartilage cells, GR1 RNA was seen in a number of other fetal skeletal tissues.

Bones of the axial skeleton form by the process of

endochondral ossification. In the area where bone is to form, mesenchymal condensations appear which outline the general shape of the bone (see Johnson, 1986 for review). The outer cells of this condsensation, the perichondrium, are elongated in parallel to the longitudinal axis of the bone, whilst the inner cells are elongated perpendicular to this axis.

As development proceeds, the perichondrium becomes bilaminar with an outer fibroblastic layer and an inner surface of relatively undifferentiated chondrogenic cells. The cells within the blastema differentiate to chondrocytes which actively synthesise and secrete cartilage matrix, i.e. chondrification occurs. Cartilage matrix is unique in that it contains large amounts of type II collagen and chondroitin sulphate proteoglycan (Stockwell, 1979). During this stage the bone increases in size by means of interstitial growth of cells within the developing matrix.

Subsequently, the inner layer of the perichondrium, now termed the periosteum, produces osteoblasts which secrete bone matrix (osteoid) resulting in the appearance of a bony layer around the cartilage and calcification of the cartilage, i.e. ossification. This period corresponds to hypertrophy and death of the chondrocyte cells. The calcified cartilage breaks down and is replaced by osteogenic cells, hemopoietic cells and capillaries. The periosteum retains the ability to produce either osteoblasts or chondroblasts and in this way appositional growth of the bone occurs. In the developed bone, chondrocytes are retained only at the epiphysal heads of the

long bones in areas of hyaline cartilage. During mouse development, the appearance of ossification centres varies from 15-18.5 days gestation, depending on the particular bone (Johnson, 1933; Rugh, 1968).

Within a given developing long bone, all these stages of development can be observed, the central regions being at later stages of differentiation than peripheral regions (Stockwell, 1979). Moreover, within a developing limb at a given point in embryogenesis, proximal bones are at a later stage of development than distal bones.

Within limb buds at 12.5 days gestation, GR1 expression was seen in the blastemal condensations of limb bud mesenchyme, and by 14.5 days gestation, expression was seen in the chondrocytes of condensed mesenchyme but not in the surrounding perichondrium (data not shown).

Ossification of humerus, radius and ulna commences at 14-15 days gestation (Johnson, 1933; Rugh, 1968). At 17.5 days gestation, GR1 expression within the humerus and radius was seen to be restricted to chondrocytes, with hypertrophoid chondrocytes and periosteum being negative for GR1 expression (fig. 5.9a,b,c). This was in contrast to the uniform expression seen within the chondrocytes of the developing hindlimb digits at 17.5 days gestation (fig. 5.9d,e,f). Ossification of hindlimb digits does not commence until 18 days gestation (Rugh, 1968), and this was reflected in the absence of hypertrophoid chondrocytes within the embryonic digit bones at 17.5 days gestation (fig. 5.9d,e,f). The

Figure 5.9. GR1 expression in developing limbs.

Longitudinal sections of 17.5 days gestation mouse embryos were prepared (section 2.19) and hybridised with sense (A,E) and anti-sense (B,C,D,F) ^{35}S radiolabelled RNA probes prepared from pBS1200A or pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours.

A and B: mid-saggital sections of elbow region of forelimbs.

C: para-sagittal section of the same region as A and B.

D,E, and F: mid-saggital sections of hindlimb digits.

A,B,C,D: 40x magnification.

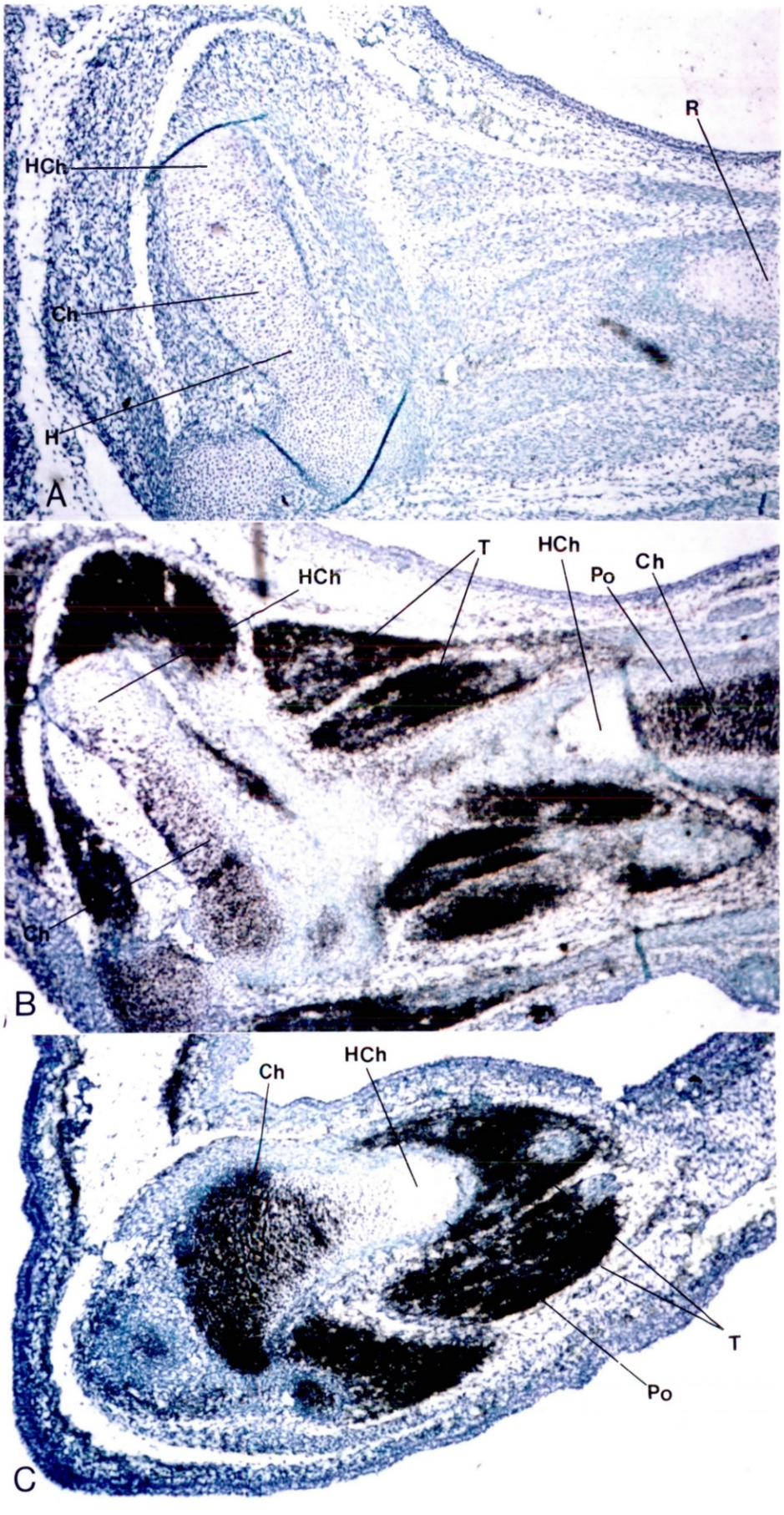
E,F: 100x magnification.

Some folding of tissues is apparent in A and D as seen by the sharp blue lines in some areas of these sections.

Ca= calcaneus, Ch= chondrocytes, Cu= cuneiforme, Ep= epidermis, H= humerus, HCh= hypertrophied chondrocytes, Mt= metatarsale, Mwt= mesodermal web tissue, Pc= perichondrium, Po= periosteum, R= radius, T= tendon, Ta= talus, Ti= tibia.

A,B and C on facing page.

D,E and F, overleaf.


R
HCh
Ch
H
A
T
HCh
Po
Ch
HCh
Ch
B
Ch
HCh
T
Po
C

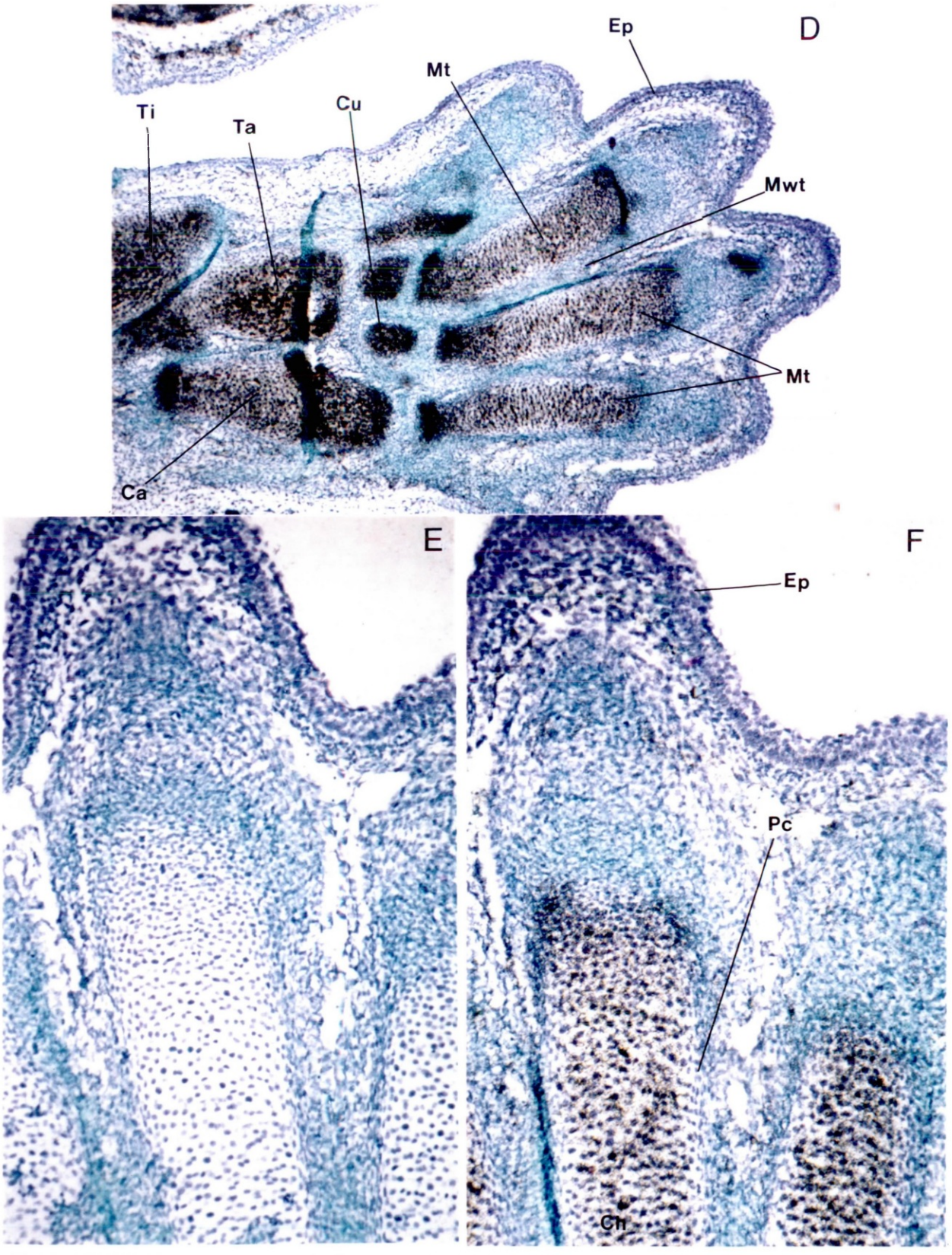

D
Ep
Mt
Cu
Ti
Ta
Mwt
Mt
Ca
E
F
Ep
Pc
Ch

Figure 5.10. GR1 expression in embryonic ribs.

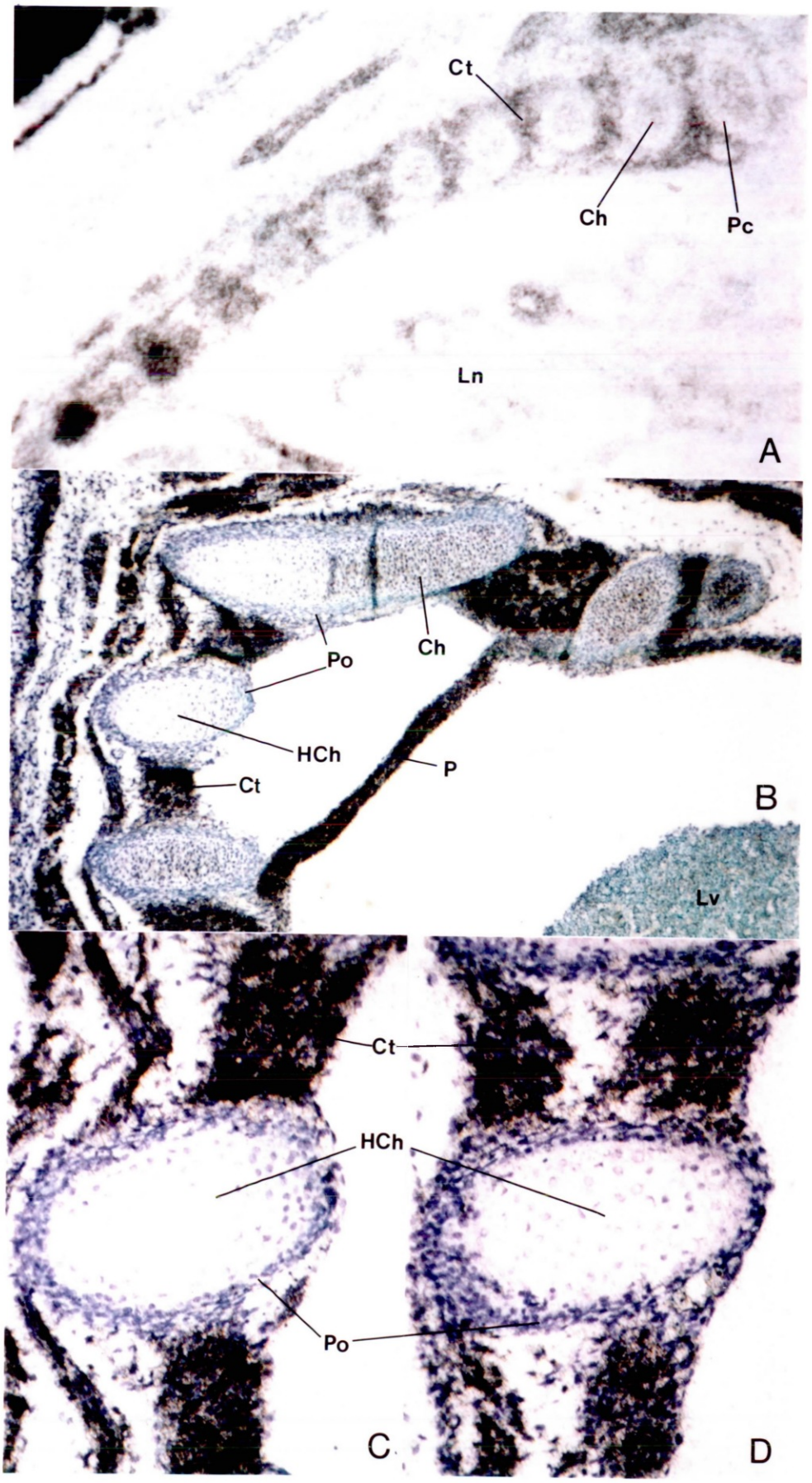
Transverse (A) and longitudinal sections (B,C,D) of 14.5 days (A), and 17.5 days (B,C,D) gestation mouse embryos were prepared (section 2.19) and hybridised with anti-sense ^{35}S radiolabelled RNA probes prepared from pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours.

A,C,D: 100x magnification.

B: 40x magnification.

Some folding of tissue in B is apparent denoted by sharp blue lines.

Ch= chondrocytes, Ct= connective tissue, HCh= hypertrophied chondrocytes, Ln= lung, Lv= liver, P= peritoneum, Pc= perichondrium, Po=periosteum.


Ct
Ch
Pc
Ln
A
Po
Ch
HCh
P
Ct
B
Lv
Ct
HCh
Po
C
D

perichondrium was seen to be negative for GR1 expression within the developing digits at this stage of development.
A similar situation was seen in the embryonic ribs where chondrification and ossification centres arise at 13 and 14 days gestation, respectively (Rugh, 1968). At both 14.5 and 17.5 days gestation, GR1 expression was seen in chondrocytes but not perichondral tissue (fig. 5.10). Similarly, in regions of ribs undergoing ossification, neither hypertrophied chondrocytes nor the periosteum were seen to express GR1 (fig. 5.10b,c,d).

5.2.4.(v) GR1 expression in lung and gut mesenchyme and other connective tissues.

The mesodermal expression of GR1, first apparent at 7.5 days gestation, persisted in the mesenchymal cells of the developing lung and gut.
Within the developing lungs at 12.5 days gestation, the bronchioles did not contain detectable levels of GR1, in contrast to the surrounding mesenchyme, in which GR1 RNA was readily detected (fig. 5.11a). A similar situation was seen in the gut at 12.5 days gestation where the surrounding mesenchymal tissue, but not the gut epithelium, expressed GR1. In both lung and gut, the condensed mesenchyme surrounding the bronchioles or gut epithelium subsequently differentiates to form connective tissue and smooth muscle.

Figure 5.11. GR1 expression in embryonic gut and lung mesenchyme.

Sections of 12.5 days gestation mouse embryos were prepared (section 2.19) and hybridised with sense (B) and anti-sense (A,C) ^{35}S radiolabelled RNA probes prepared from pBS1200A or pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours.

All at 100x magnification.

B= bronchioles, GE= gut epithelium, GM= gut mesenchyme, LM= lung mesenchyme.

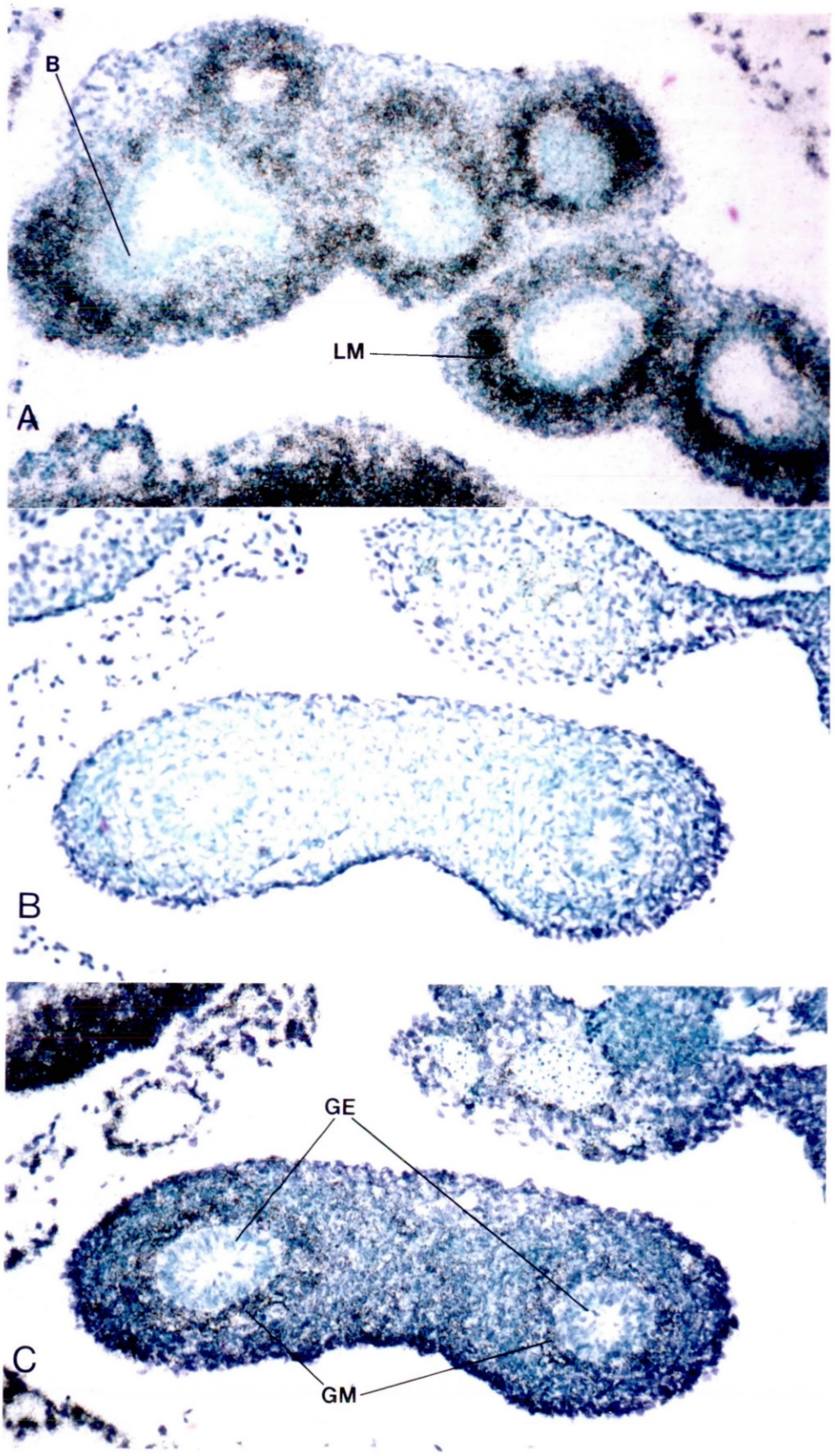
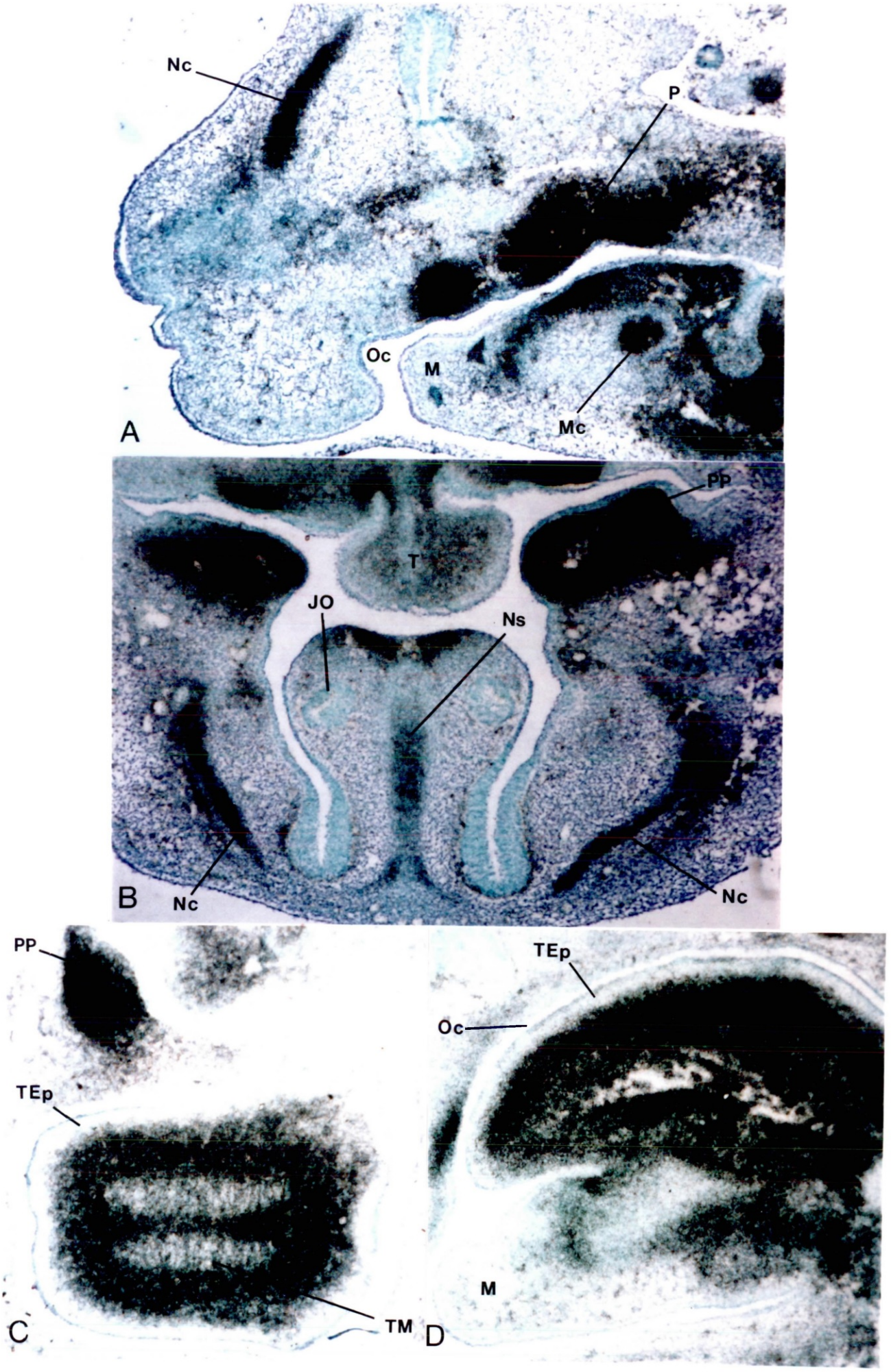

B
LM
A
B
GE
GM
C

Figure 5.12. GR1 expression in developing craniofacial bones and connective tissues.

Longitudinal (A,D) and transverse (B,C) sections of 12.5 days gestation mouse embryos were prepared (section 2.19) and hybridised with anti-sense ^{35}S radiolabelled RNA probes prepared from pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours. All at 100x magnification.

JO= Jacobson's organ, M= mandible, Mc= Meckel's cartilage, Nc= nasal capsule, Ns= nasal septum, Oc= oral cavity, P= palate, PP= palatine process, T= tongue, TEp= tongue epithelium, TM= tongue muscle.


Nc
P
Oc
M
Mc
A
PP
T
JO
Ns
B
Nc
Nc
PP
TEp
Oc
TEp
M
C
TM
D

GR1 expression was also detected in the connective tissue between ribs at 14.5 days and 17.5 days gestation (fig. 5.10), tendons in limbs (fig. 5.9), the peritoneum (fig. 5.10) and liver (fig. 5.15c). Within the tongue, expression was seen in the developing connective tissues but not in the overlying epithelium or inner muscle blocks (fig. 5.12).

The association of GR1 expression with developing skeletal structures was also reflected in the expression observed within the nasal septum and capsule, Meckel's cartilage, palatine process, and both upper and lower palates (fig. 5.11), as well as hyoid cartilage and other developing craniofacial bones (data not shown).

5.2.4.(vi) GR1 expression in developing neural tissues.

At 9.5 days gestation, GR1 RNA was seen to be localised to discrete regions of the mesencephalon and myelencephalon, and expression was seen in head mesenchyme (fig. 5.6a). At later stages of development, GR1 expression was maintained in specific regions of the brain and spinal cord.

At 12.5 days gestation, GR1 RNA was detected at the base of the optic stalk (fig. 5.13a) and by 17.5 days gestation, GR1-expressing cells were seen to lie around the optic nerve (fig. 5.13b). This may be attributable to the expression seen within developing cranial bones and connective tissue. However, GR1 expression in obvious neural tissue was also observed.

Figure 5.13. Localisation of GR1 RNA in fetal brain and spinal cord.

Transverse sections of 11.5 days (A), 14.5 days (C) and 17.5 days (B) gestation mouse embryos were prepared (section 2.19) and hybridised with anti-sense ^{35}S radiolabelled RNA probes prepared from pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours. Some fragmentation of lens tissue at 17.5 days gestation is apparent in B. Dark staining around the lens in B is due to the pigment layer of the retina and not hybridisation of radiolabelled probe.

A,C: 100x magnification.

B: 40x magnification.

Ls= lens, Lsv= lens vesicle, No= notochord, ON= optic nerve, OS= optic stalk, Pl= pigment layer of retina, R= retinal layer, SC= spinal cord, VC= vertebral cartilage.

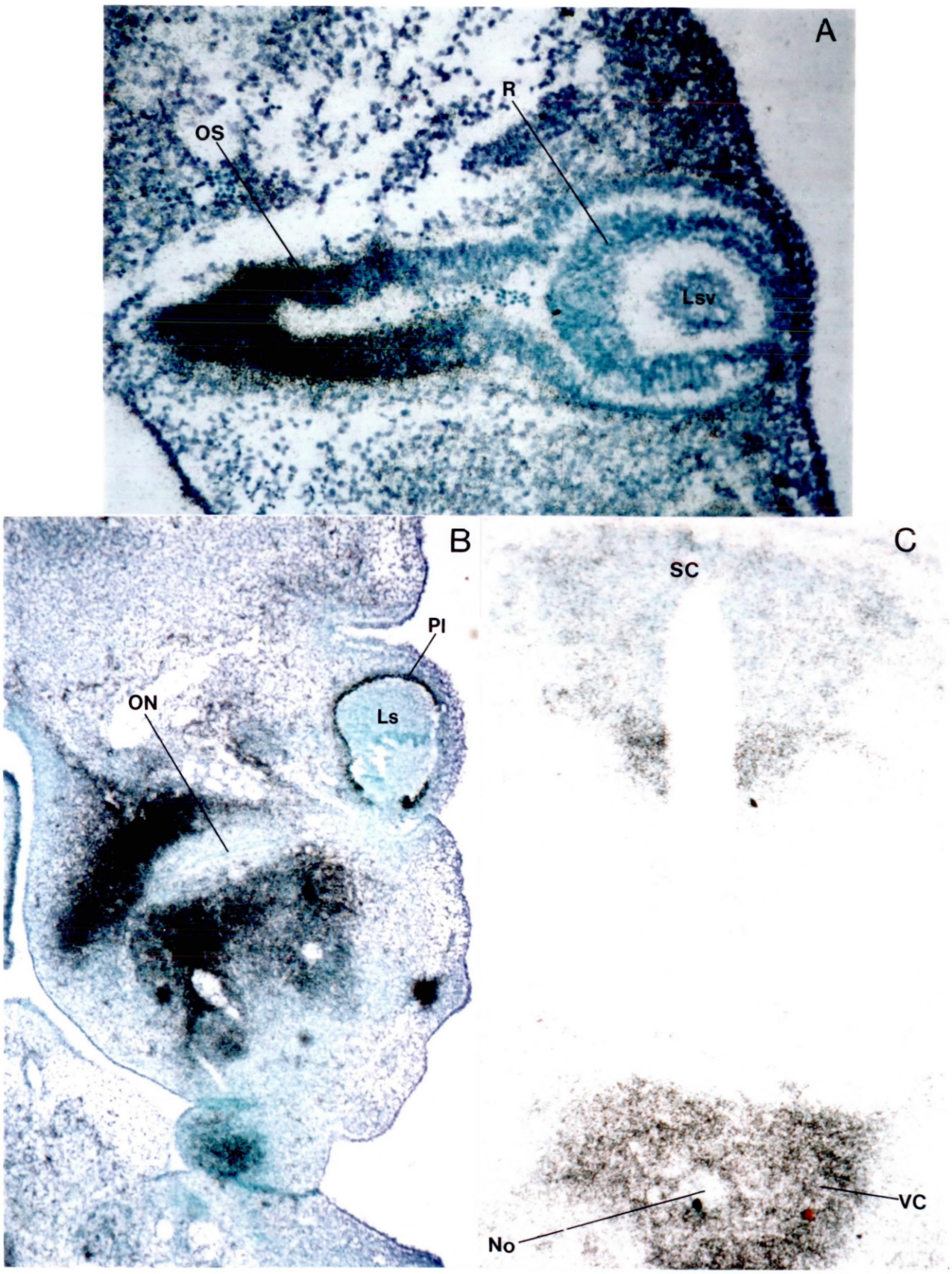

A
OS
R
Lsv
B
Pl
ON
Ls
C
SC
VC
No

Figure 5.14. GR1 expression in developing pituitary gland.

Serial transverse sections of a 12.5 days gestation mouse embryo were prepared (section 2.19) and hybridised with anti-sense ^{35}S radiolabelled RNA probes prepared from pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours.

A. Relationship of Rathke's pocket and infundibulum in the mid-gestation mouse embryo.

B and C represent the same section but at different magnifications. B/C, D and E represent a caudal to cranial series of sections.

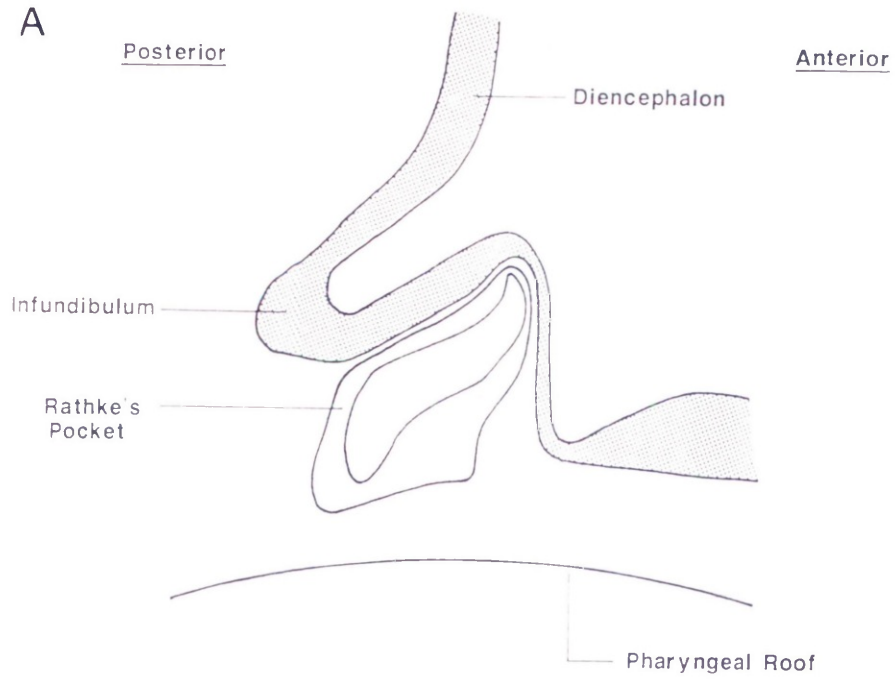
B: 40x magnification.

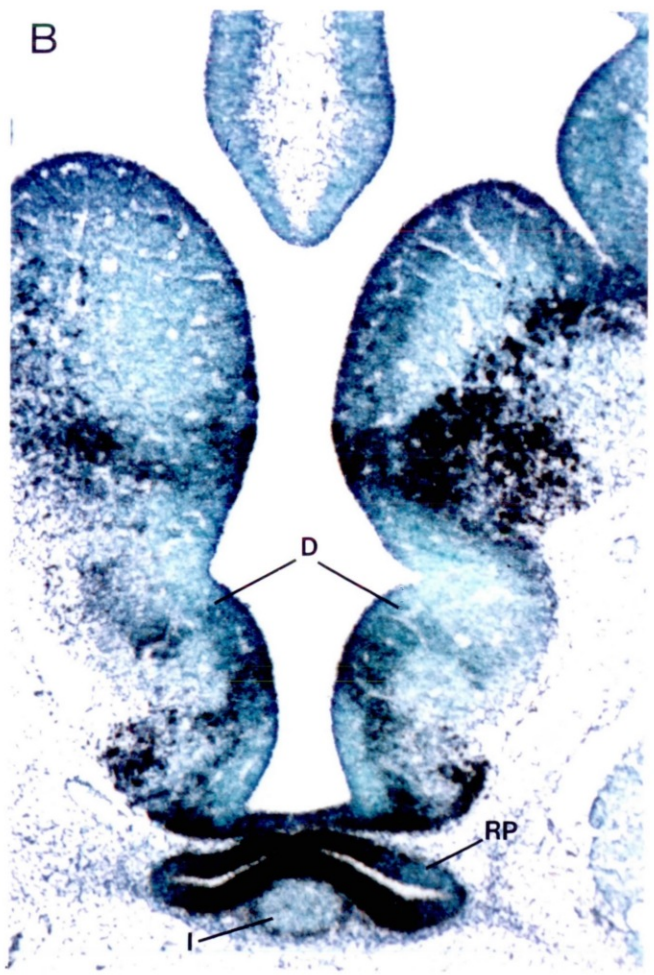
C,D,E: 100x magnification.

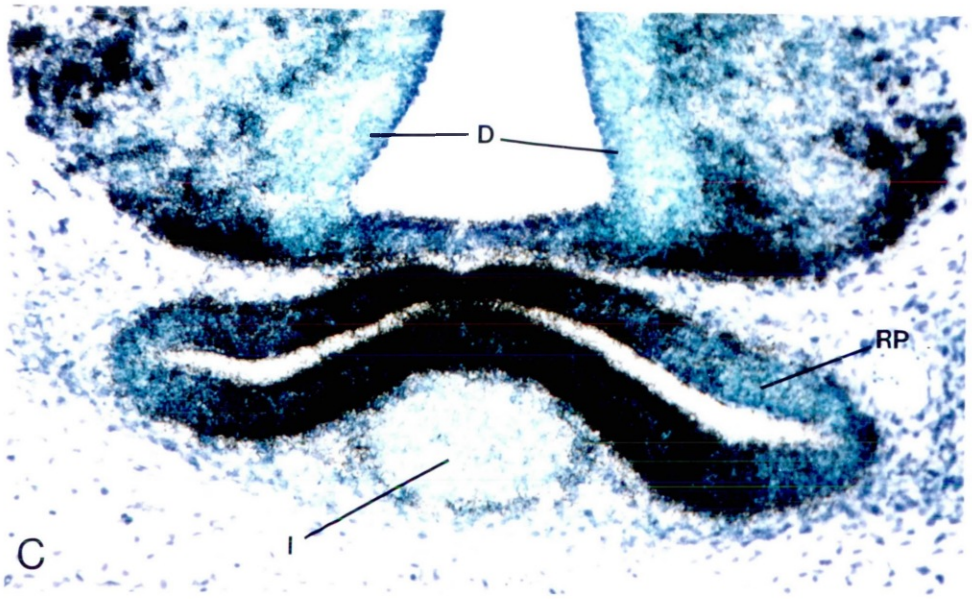
D= diencephalon, I= infundibulum, RP= Rathke's pocket.

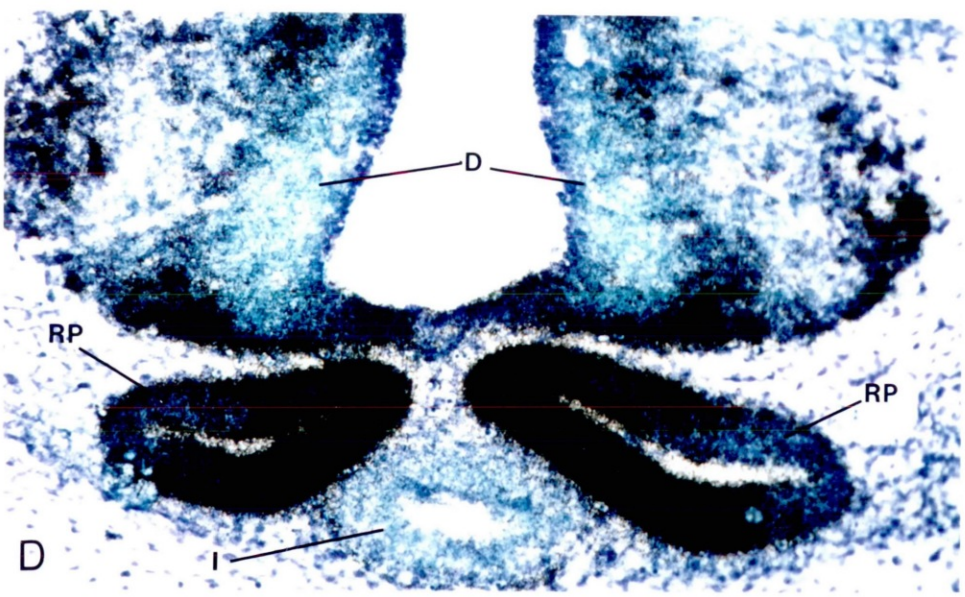
A and B facing page.

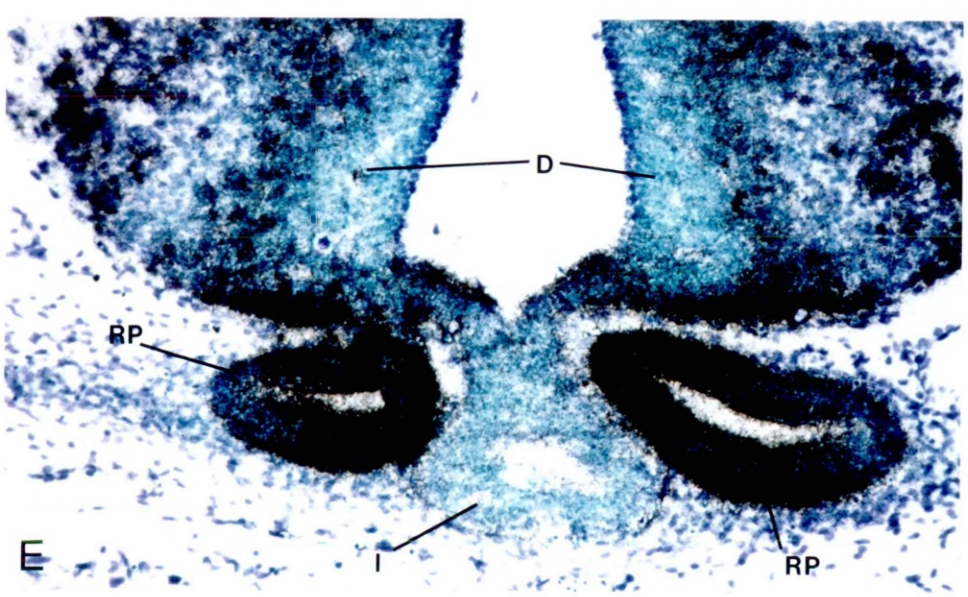
C,D and E, overleaf.


A
Posterior
Anterior
Diencephalon
Infundibulum
Rathke's
Pocket
Pharyngeal Roof


B
D
RP
I


D
RP
I
C


D
RP
RP
I
D


D
RP
I
RP
E

At 14.5 days gestation, a specific band of GR1-expressing cells was apparent in transverse sections of spinal cord (fig. 5.13c), whilst specific regions of the diencephalon were also seen to contain GR1-expressing cells at 12.5 days gestation (fig. 5.14a). However, the most striking expression pattern of GR1 within the developing brain was that in the pituitary gland (fig. 5.14).

The pituitary gland develops from an upgrowth of the roof palate overlying the oral cavity, Rathke's pocket, and a downgrowth of the diencephalon, the infundibulum (Theiler, 1972; fig. 5.14a). Rathke's pocket comes to surround the infundibulum both below, anteriorly, and on either side, leaving only a thin contact between the infundibulum and the diencephalon. Rathke's pocket forms the pars anterior, pars intermedia and pars distalis endocrine regions of the pituitary, whilst the infundibulum forms the pars nervosa, the neural region of the pituitary (Theiler, 1972). Being composed of both endocrine and neural tissues, the pituitary serves as a link between the nervous system and endocrine organs.

From the sequence of serial sections shown in fig. 5.14, it is apparent that particularly high levels of GR1 were seen in Rathke's pocket at 12.5 days gestation. Such expression was observed from the first appearance of this tissue at 9.5 days gestation (data not shown). Moreover, a localisation of GR1 expression towards the central region of Rathke's pocket was observed, along with lower levels of expression in cells immediately surrounding the infundibulum (fig. 5.14). The

infundibulum itself did not express GR1 at 12.5 days gestation (fig. 5.14). Later stages of pituitary gland development were not investigated.

5.2.4.(vii) Localisation of GR1 RNA in embryonic and adult testis.

Within the 10.5 days gestation mesonephros, GR1 expression was detected in the mesonephric mesenchyme but not in the mesonephric tubules (fig. 5.15a,b).

In the developing testis at 12.5-13 days gestation, GR1 RNA was readily detected in the small interstitial cells of the gonad stroma around the future seminiferous tubules, particularly those interstitial cells in the centre of the organ (fig. 5.15c,d). Within the tubules, the developing spermatogenic cells, derived from primordial germ cells, were found not to express GR1. Similarly, the limiting epithelium of the developing testis did not contain detectable levels of GR1 RNA (fig. 5.15d).

Low levels of GR1 RNA were detected in sections of adult mouse testis (fig. 5.15e,f). As in the embryo, expression was seen to be restricted to the non-spermatogenic cells at the periphery of the seminiferous tubules. No GR1 RNA was detected in the spermatogenic cells within the seminiferous tubules (fig. 5.15e,f; data not shown).

Figure 5.15. Localisation of GR1 RNA in embryonic and adult mouse testis.

Sections of 10.5 days (A,B) and 12.5 days (C,D) gestation embryos, along with sections of adult mouse testis (E,F), were prepared (section 2.19) and hybridised with sense (A,E) and anti-sense (B,C,D,F) ^{35}S radiolabelled RNA probes prepared from pBS1200A and pBS1200B as described in section 2.20.5. Autoradiography (section 2.21.3) was carried out for 48 hours. Some fragmentation of liver tissue was apparent in C.

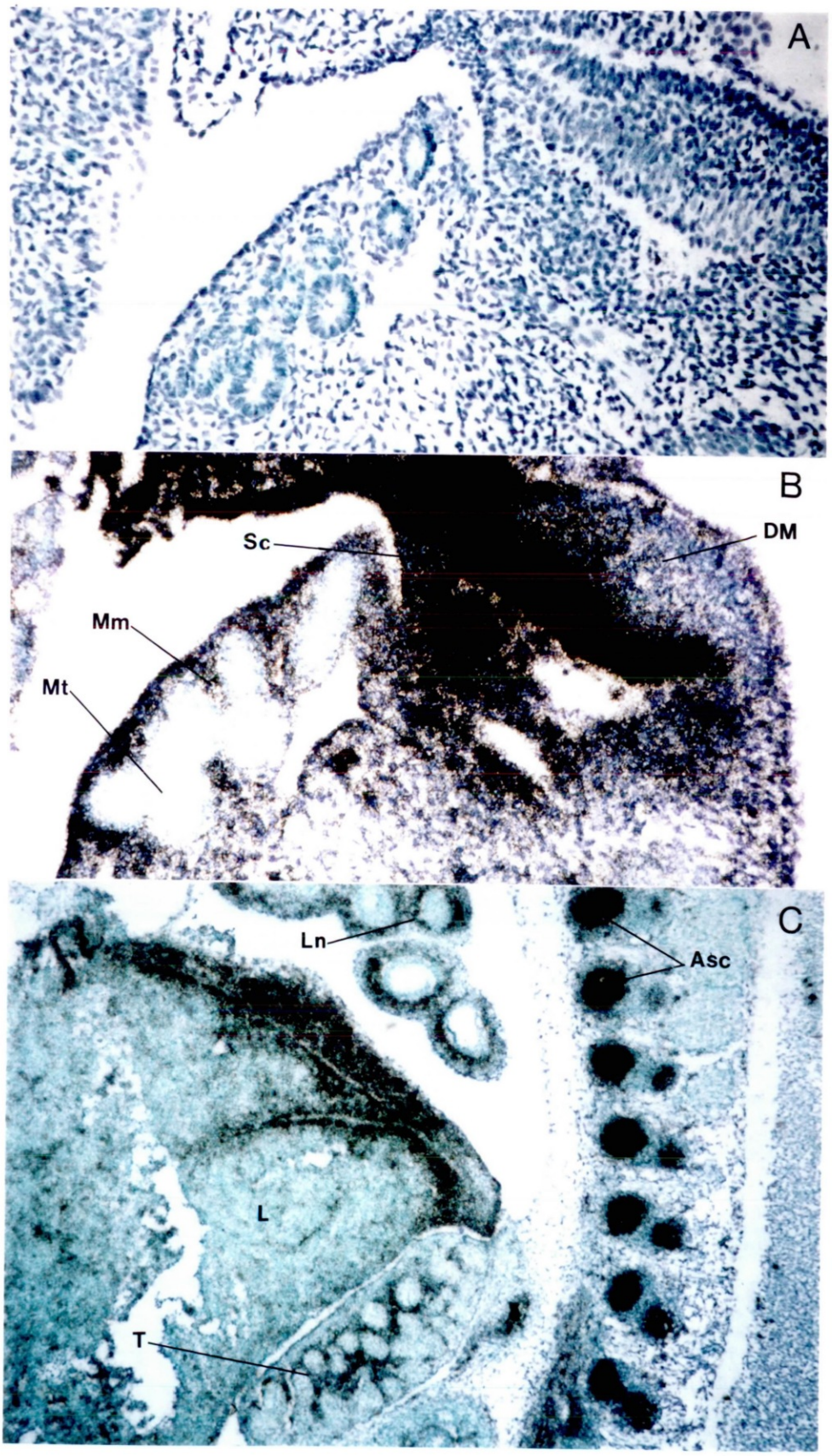
C: 40x magnification

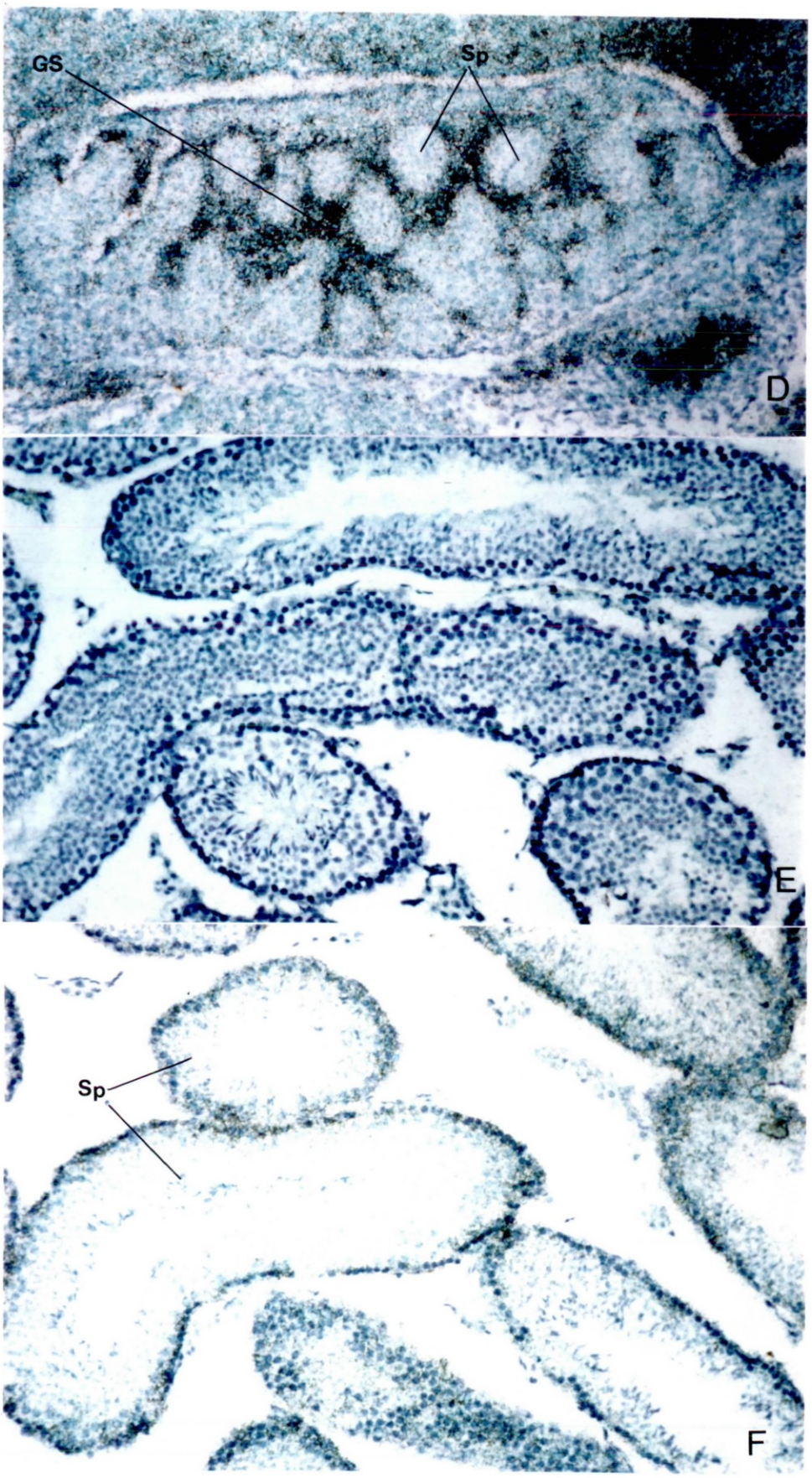
A,B,D,E,F: 100x magnification.

Asc= anterior sclerotome, DM= dermamyotome, GS= gonad stroma, L= liver, Ln= lung, Mm= mesonephric mesenchyme, Mt= mesonephric tubules, Sc= sclerotome, Sp= spermatogenic cells, T= testis.

A,B and C on facing page.

D,E and F, overleaf.


A
B
Sc
DM
Mm
Mt
C
Ln
Asc
L
T


GS
Sp
D
E
Sp
F

Section 5.3. Discussion.

Section 5.3.1. GR1-fusion protein studies.

The availability of antisera against GR1-encoded antigens would greatly facilitate the analysis and characterisation of the expression pattern, properties and potential functions of this novel gene. In the absence of any nucleotide sequence data, the production of β-galactosidase fusion proteins in E.coli, using the cDNA clone pGR165, provided a potentially rapid route to the production of GR1-specific polyclonal antisera.

Initial attempts to generate fusion proteins using the entire 2.5kb cDNA fragment of pGR165 proved unsuccessful. The absence of clones containing the sense orientation of the cDNA in any of the three pUR vectors suggests that the problem was one of DNA rearrangements within the recombinant plasmids rather than the synthesis of a toxic fusion protein. The apparent instability of clones containing the cDNA restriction fragment in one orientation only could be due to the occurrence of regions of DNA sequence homology between the plasmid vector and the cDNA restriction fragment. When cloned in one orientation, in this case anti-sense, these elements would be arranged in a relatively stable tandem formation. However, in the other orientation these elements would now be inverted relative to each other. As discussed earlier (section 4.1.3), palindromic sequences are often unstable in rec^+ $sbcB^+$ host

E.coli strains. As with genomic clones carrying palindromic sequences, propagation of clones under non-recombinogenic conditions may facilitate the synthesis of β-galactosidase-GR1 fusion proteins synthesised from the entire cDNA restriction fragment of pGR165.

The 1.1kb PstI-PstI restriction fragment of pGR165 proved to be stable, in both orientations, in each of the pUR vectors used. The sense strand of this fragment, when cloned 3′ to the lacZ gene, was shown to contain a single large open reading frame which, in the clone 291-12, was in the same frame as the lacZ gene. The detection of this protein in three independent clones strongly suggested that this fusion protein was not a cloning artefact. The size of this fusion protein (121kd) indicates that this fragment of pGR165 contains a single open reading frame of approximately 150 nucleotides from the 5′ PstI site of the restriction fragment used for fusion protein analysis. Thus, the cDNA clone pGR165 is predicted to contain a single open reading frame of approximately 650 nucleotides, including the repetitive element (section 4.2.2), the remainder of the clone representing 3′ untranslated sequences.

Immunisation of rabbits with the 291-12 fusion protein resulted in the production of antisera against this fusion protein. On the assumption that the presence of GR1 mRNA correlates with the presence of GR1 antigens, Northern blot data (section 3.2.7) predict that GR1 antigens are present in PCC3 cells but absent, or present at markedly reduced levels,

in retinoic acid treated PCC3 cells and the SV40-transformed cell line SV3T3 C138. However, the 291-12 fusion protein antiserum was found to lack any activity against antigens which showed this distribution of expression. Such negative results could be attributed to one or more of a number of factors.

The GR1 polypeptide domain represented in 291-12 may fold differently when expressed as a β-galactosidase fusion protein than when part of the complete GR1 polypeptide. Also, the GR1 domain represented in 291-12 may correspond to a poor antigenic site, such as a highly hydrophobic protein domain. Alternatively, the 291-12 fusion protein may be artefactual. It may represent an open reading frame in the GR1 mRNA which is not translated in vivo. Copying errors incorporated into pGR165 during cDNA synthesis may have given rise to an open reading frame not found in the GR1 mRNA. It remains formally possible that one of the other two reading frames of the sense strand of the restriction fragment used represents the true open reading frame of pGR165. This open reading frame may not be of sufficient size to show a detectable change in migration rate on SDS-PAGE. Also, fusion proteins containing the open reading frame used in mammalian cells may be highly unstable in E.coli, being degraded to a stable form of size similar to that of native β-galactosidase. It is formally possible that GR1 RNA does not encode a protein in PCC3 cells, being a structural RNA, or encodes a protein which is rapidly turned over within PCC3 cells.

Whatever the reason, it was apparent that no anti-GR1 serum had been obtained. Further attempts to raise antisera against any putative GR1 polypeptide(s) were deferred until the complete DNA sequence of pGR165, as well as DNA sequence of coding regions of the GR1 genomic clone N28, were available. Analysis of such sequence data (section 6.2.4) indicated that the open reading frame utilised in 291-12 lies in the 3′ untranslated region of the GR1 transcription unit.

Section 5.3.2. GR1 expression during mouse embryogenesis.

From the data presented in figures 5.3 to 5.15, it is clear that GR1 is expressed at a number of discrete sites during mouse development. Whilst the results do not represent an exhaustive analysis of all cell types at all stages of embryogenesis, they do serve a useful purpose in illustrating at least some of the sites of expression of this novel gene and can be used as a basis for more detailed expression studies.

5.3.2.(i) Relating GR1 expression in the early embryo to embryo-derived stem cell lines.

From analysis of 7.5 days gestation conceptus, it was seen that whilst expressed at high levels in extraembryonic

tissues, and at lower levels in mesodermal cells arising as the primitive streak, there was little, or no, GR1 expression in primitive ectoderm cells. This result is sharp contrast to the expression seen in both EC and ES cell lines (section 3.2).

As discussed earlier (section 1.2), both EC and ES cells represent pluripotent cell types within the early mouse embryo. The EC cells used in this work were derived from a teratocarcinoma, OTT6050, produced by transplantation of a 6 days gestation, egg cylinder stage, mouse embryo to an ectopic site (Stevens, 1970). Similar experiments indicated that transplantation of 7.5 days gestation embryos to ectopic sites failed to give rise to teratocarcinomas containing cells of EC morphology (Stevens, 1970). This, along with evidence from *in vitro* culture of 7 days gestation egg cylinders (Snow, 1981; see Snow, 1985 for review), suggests that, between 6 days and 7.5 days of mouse development, there is a restriction in the developmental capacity of embryonic ectoderm.

Given that GR1 RNA is readily detectable in pluripotent EC cells, and ES cells derived from early mouse blastocysts (section 3.2), it is predicted that GR1 expression should be apparent within the pluripotential cells of the inner cell mass and early embryonic ectoderm of the mouse embryo. Expression would then be lost in the late egg cylinder embryonic ectoderm only to reappear in the mesodermal cells of the primitive streak during gastrulation. Additional RNA *in situ* hybridisation analyses of early mouse embryos are

required to test this prediction. In addition, such analyses will determine the point in embryogenesis at which GR1 expression in trophectoderm, and other extraembryonic tissues, is induced.

It is also possible that the expression of GR1 in EC and ES cells is simply an artefact of *in vitro* culture of murine embryonic stem cells. Within the blastocyst and egg cylinder, the pluripotential embryonic stem cells are components of a 3-dimensional structure containing a variety of cell types, in sharp contrast to growth as clonal monolayers *in vitro*. Moreover, in the establishment and maintenance of EC cell lines, there has been selection both *in vivo* and *in vitro* for EC cells which will not undergo spontaneous differentiation. The analyses of GR1 expression within mouse embryos at stages earlier than 7.5 days gestation, discussed above, will help to clarify whether GR1 expression seen in embryo-derived stem cells lines is an artefact of *in vitro* culture of such cells.

5.3.2.(ii) GR1 expression during somitogenesis and bone development.

Having established that GR1 is expressed in mesenchymal cells arising in the primitive streak, it was seen that expression was maintained in a number of mesenchymal-derived tissues. The general expression in presomitic mesenchyme of the tail bud at 9.5 and 10.5 days gestation was in striking contrast to

the clear restriction of GR1 expression seen within the differentiated somite. Specific mechanisms must operate within the somite in order that GR1 expression is maintained in sclerotome and yet lost in the dermamyotome of the somite. By 12.5 days gestation, a second restriction to GR1 expression was observed within the sclerotome. Expression was specifically maintained in the anterior sclerotome, the densely packed posterior sclerotome half containing no detectable GR1 RNA. Such a pattern of expression is interesting in relation to the known properties of sclerotome halves in vertebrates.

From grafting experiments using chick/quail chimeras, it has been established that motor and sensory neurones emerging from the neural tube will only grow through the anterior sclerotome half (Keynes and Stern, 1984). Similarly, the migration of neural crest cells through somites occurs only through the anterior somite half (Keynes and Stern, 1985, 1986; Rickmann et al., 1985; Bronner-Fraser, 1986). Migration of neural crest cells occurs prior to axon growth, but axon growth has been shown to be independent of neural crest cell migration (Rickmann et al., 1985). It has been established that such commitment to anterior or posterior halves is established immediately after segmentation of the somites, prior to sclerotome formation (Stern and Keynes, 1986).

At the molecular level, peanut agglutinin has been found to bind specifically to the posterior half as soon as sclerotome forms (Stern et al., 1986), in contrast to laminin and

fibronectin which are evenly distributed about both sclerotome halves (Rickmann et al., 1985). The neural crest cell antigen HNK-1 is evenly distributed across the sclerotome during neural crest cell migration (Rickmann et al., 1985), as are glycosaminoglycans during this period (cited in Keynes and Stern, 1986). In the mouse, it has been shown that expression of a murine paired-box gene, PAX-1, localises specifically to the posterior sclerotome half in 12.5 days gestation mouse embryo (Deutsch et al., 1988). Such an expression pattern is complementary to that of GR1 which localises to the anterior sclerotome half at this stage of development.

It will be of interest to determine whether the distribution of GR1 protein correlates with that of GR1 RNA in this tissue. Moreover, given the suitability of the chick embryo in the analysis of vertebrate somitogenesis and neural development, it would be useful to investigate the expression pattern of any chick homologue of GR1.

GR1 expression was seen to be maintained within the vertebral cartilage cells at 14.5 days gestation, whilst the developing intervertebral disc did not express this gene. As at 12.5 days gestation, this expression pattern was complementary to that seen for PAX-1 (Deutsch et al., 1988). Unlike many murine homeobox-containing genes (section 5.1.2), no localisation of GR1 expression was observed within somites along the anterior-posterior axis of the embryo, it being expressed at equivalent levels in all sclerotomes analysed, although expression in cranial somites was not investigated.

GR1 expression at later stages of vertebrae development was not investigated, but it would clearly be of interest to determine whether expression of this gene is maintained in ossified vertebrae or is restricted to the chondrified stage of vertebrae development. Such a restriction to the cartilaginous stage, and specifically chondrocytes, of developing bones was suggested by analysis of GR1 expression within embryonic limbs and ribs.

Ribs are derived from thoracic sclerotome, known to express GR1, whilst limb buds develop from the lateral mesoderm plate. From analysis of both ribs and long bones of the limbs, it was seen that GR1 was expressed specifically in the transient chondrocyte population which forms the cartilage model for bone development. GR1 RNA was not detected in the overlying perichondrium, nor in hypertrophied chondrocytes or periosteum at later stages of embryogenesis when ribs and proximal limb bones are known to be undergoing ossification (Rugh, 1968; Theiler, 1972). This restriction to the chondrocytes of developing long bones is in contrast to c-fos expression in embryonic bone, where expression was observed in the perichondral growth regions and mesodermal web tissue of developing limbs at 17 days gestation, but not in condensed limb mesenchyme at either 10 days or 13 days gestation (Dony and Gruss, 1987b). Also, in contrast to GR1, osteonectin has been shown to be expressed in periosteum and osteoblasts, but not cartilaginous tissues, in the long bones of newborn mice (Holland et al., 1987).

Confirmation that GR1 expression within embryonic skeletal structures is limited to the chondrogenic, and not osteogenic, stages of bone development could come from further RNA in situ hybridisation analyses in which serial tissue sections were probed for GR1 expression and stained with Alizarin Red S, a stain specific for ossification centres (Rugh, 1968).

It remains to be established whether GR1 expression persists in cartilaginous tissues of the adult mouse or is restricted specifically to the chondrocyte population within developing bones. Additional Northern blot and RNA in situ hybridisation experiments will help clarify this situation. Further analyses are also required to determine the temporal restriction of GR1 expression within the limb bud mesenchyme during embryogenesis.

5.3.2.(iii) GR1 expression in connective tissues and other mesenchyme derivatives.

In addition to expression within developing bones, GR1 RNA found to be associated with a number of connective tissue types, including that between ribs, the peritoneum, tendons, and gut and lung mesenchyme. In the latter cases, a clear differential pattern of expression was observed, with neither the gut epithelium nor bronchioles containing detectable levels of GR1 RNA.

Differential expression between mesenchymal and epithelial

cells was also apparent in the developing mesonephros. At 10.5 days gestation the mesonephric tubules did not express this gene, in contrast to the surrounding mesenchyme. GR1 expression was also apparent in the gonad stroma, whereas the developing seminiferous tubules and spermatogenic cells did not express this gene. Such a restriction of GR1 expression to interstitial cells of the testis was maintained in the adult organ.

The origin of gonadal stromal cells is unclear, it being suggested that they arise from coelomic epithelium (Callebaut 1976), mesonephric blastema (Carlon et al., 1983) or from differentiated mesonephros (reviewed by Rodemer et al., 1986). That the mesonephros participates in the generation of gonad stroma in the chick has been established from chick/quail chimera experiments (Rodemer et al., 1986) and there is also evidence for a mesonephric origin of the gonad somatic cells in the mouse (Upadhyay et al., 1979).

Whilst GR1 expression within the testis has been investigated, any expression within the ovaries is yet to be established. Given the data presented here, it would be of interest to determine whether GR1 was expressed in the somatic cells of embryonic and mature female gonads or was restricted only to those of the male.

5.3.2.(iv) Craniofacial structures and GR1 expression.

The expression of GR1 within craniofacial skeletal structures, and tongue connective tissues, is compatible with the observed expression of this gene in chondrocyte populations of developing vertebrae, limbs and ribs, and connective tissues in other regions of the embryo.

However, it should be noted that the ossification of craniofacial bones occurs by membranous ossification, a process distinct from the endochondral ossification of the axial skeleton. Although both processes involve mesenchymal condensations, in membranous ossification the mesenchyme forms fibrocellular, or membranous, tissue within which ossification occurs (Gardner, 1971). In the mouse mandible, whilst chondrocytes undergo hypertrophy they do not, unlike endochondral ossification, die but contribute to bone formation (Silbermann and Frommer, 1972). More detailed RNA <u>in situ</u> hybridisation analyses are required in order to establish the precise cell types within craniofacial structures that express GR1.

In addition to the process of ossification, the origin of craniofacial structures, which contain GR1 RNA, is different from those of other skeletal elements which express this gene. Craniofacial skeletal structures are derived from both mesenchyme of paraxial mesoderm and ectomesenchyme derivatives of neural crest cells.

Neural crest cells have their origin in neural folds and the

neural plate, from which they migrate to precise locations in vertebrates and differentiate into a variety of cell types, including those of the peripheral nervous system, adrenal medulla and both skeletal and connective tissues of craniofacial structures (Hall, 1982; Le Douarin, 1982; Noden, 1983a,b; D'Amico-Martel and Noden, 1983). Whilst in the chick neural crest migration follows neural tube closure, in the rat migration is seen both during and after neural tube formation (Tan and Morriss-Kay, 1985) and neural crest cells have been observed to migrate through head mesenchyme in a diffuse pattern (Tan and Morriss-Kay, 1986). At the start of migration, neural crest cells are pluripotential but become restricted as a result of environmental influences during migration and at the site of arrest. They appear in both cranial and trunk regions of the neural tube.

Much of the knowledge concerning the origin, migration and derivatives of neural crest cells in vertebrates has come from anlayses of chick/quail chimeras. From such experiments, it has been established that connective tissue of the lower jaw and tongue is of neural crest cell origin (Le Lievre and Le Douarin, 1975). Moreover, mesenchyme of the maxillary process is derived from mesencephalic neural crest and nasal process is derived from prosencephalic neural crest, whereas the cranial vault is of mesodermal origin (Le Lievre, 1978). Some tissues, such as mandibular and hyoid bones, and palatine processes are of mixed mesodermal/neural crest cell origin (Le Lievre, 1978). In mammals, a contribution of neural crest to

craniofacial structures has also been shown (Tan and Morriss-Kay, 1986).

Given that GR1 was observed to be expressed in crest cell derivatives, it would be of interest to investigate the expression of this gene within the progenitor neural crest cells and other derivatives of this migratory cell population. In this respect, the GR1 expression seen in certain areas of neural ectoderm at 8.5-9.5 days gestation may be attributable to neural crest cell types. Whilst the dorsal root ganglia of the trunk were found not to express GR1 at 12.5 days gestation (fig. 5.8), further analyses of neural crest cells, and their various derivatives, such as the adrenal medulla, are required in order to determine the temporal and cell-type specificity of GR1 expression in this migratory cell population. The ability to culture and differentiate mouse neural crest cells in vitro (Ito and Takeuchi, 1984) may provide a useful system to complement further embryological analyses.

Since some craniofacial structures are of mesodermal origin, it is also important to analyse GR1 expression within the cranial mesoderm. Such analyses would also be of interest with regard to the correlation between GR1 expression in anterior sclerotome halves and axon growth (section 5.3.2 (ii)). It is known that ganglia do not develop in regions of cranial (occipital) somites. In the chick it has been shown that, as in the trunk, crest cells will migrate into the anterior sclerotome half of cranial somites but, in contrast, no sensory ganglia develop in cranial regions (Lim et al.,

1987). Thus, it was concluded that the anterior sclerotome halves of occipital somites lack a factor essential for ganglion development (Lim et al., 1987). It would be interesting to determine whether GR1 expression in cranial somites is similar to that seen in the sclerotome of caudal somites during mouse embryogenesis.

The localised expression pattern of GR1 within the Rathke's pocket component of the developing pituitary gland is particularly striking. Further RNA in situ hybridisation analyses are required in order to establish whether this expression is specific to the developing pituitary or persists in the adult organ. The localisation of GR1 expression to areas of Rathke's pocket associated with the neural-derived infundibulum, and localised expression within other regions of the developing brain, strengthens the correlation seen between GR1-expressing cells and neural development, suggested initially by GR1 localisation to anterior sclerotome halves. Clearly, additional analyses are required in order to test this correlation. It will be particularly important to determine whether GR1 protein distribution across the sclerotome is the same as that observed for GR1 RNA.

Section 5.3.3. A role for GR1 in mouse development?

From the determination of the sites of GR1 expression during mouse embryogenesis, it was hoped to gain some indication of potential functions of this gene. Clearly, the data presented here provide only a preliminary analysis of GR1 expression in the embryo and, as such, only limited speculation can be made concerning potential functions of this novel gene.

A common feature of extraembryonic tissues, embryonic connective tissues, chondrocytes, membranes and interstitial stroma, is that they all represent structural tissues, the cells of which often secrete a variety of extracellular proteins which together constitute an extracellular matrix. This matrix is responsible for much of the structural and/or protective functions of such tissues. Extracellular matrices have also been shown to play critical roles in cell-cell and cell-substratum interactions, particularly epithelial-mesenchymal interactions during vertebrate development (see Brinkley and Morris-Wiman, 1984; Hall, 1986; Lash and Yamada, 1986 for reviews). In this respect, it is interesting to note the clear differential expression of GR1 between mesenchymal cell types and associated epithelia in embryonic gut, lung, tongue and gonad.

Some of the components of both cartilage and basement membrane extracellular matrices have been characterised at the biochemical level (see Heinegard and Paulsson, 1987; Timpl et al., 1987 for reviews). Two major components of such matrices

are collagens and proteoglycans. Both classes of components represent diverse protein families, diversity being generated both by a number of related structural genes and various post-translational modifications.

Whilst related, each of the various collagen types of vertebrates shows characteristic tissue distribution patterns (see Bornstein and Sage, 1980; Cheah, 1985 for reviews). RNA in situ hybridisation analysis, in agreement with previous expression data (reviewed by Cheah, 1985), has shown that, unlike the α1(II) collagen gene, collagen types I and III are not expressed in the chondrocytes of developing human long bones (Sandberg and Vuorio, 1987). Thus, it is unlikely that GR1 represents collagen types I or III. Nucleic acid and amino acid sequence data are available for a large number of the known interstitial and short chain collagens (section 7.2.1). Comparison of GR1 sequence data with such collagen sequences would soon establish any relationship between GR1 and this gene family.

Proteoglycans are complex glycoconjugates in which one or more glycosaminoglycan chains are covalently linked to a core protein. The high proportion of carbohydrate, often 80-90% of total mass, and large protein size has frustrated attempts to characterise proteoglycan core proteins. However, 10 or so proteoglycans have been purified, and a few core protein genes have been cloned (see Hassell et al., 1986 for review; section 7.2.1). Large and small cartilage proteoglycans, along with cell surface and basement membrane

proteoglycans have been characterised (see Hassell et al., 1986 for review).

Both collagens and proteoglycan core proteins contain characteristic repetitive amino acid sequences (Bornstein and Sage, 1980; Cheah, 1985; Hassell et al., 1986). Thus, in addition to establishing whether GR1 corresponds to a previously characterised gene, nucleotide sequence determination of the GR1 transcription unit may indicate that GR1 represents a novel member of collagen or proteoglycan core protein gene families.

It should be emphasised that the analyses presented here indicate only the distribution of GR1 RNA during mouse development. Clearly, there is a need for GR1-specific antisera in order to determine whether the distribution of GR1 protein(s) correlates with the RNA distribution. Moreover, such antisera are essential in order to determine whether GR1 encodes an intracellular or extracellular protein.

As discussed earlier (section 5.3.1), nucleotide sequence analysis would also facilitate the design of experiments to raise GR1-specific antisera. Thus, further analyses and speculation concerning putative function(s) of a GR1 protein were deferred until the amino acid sequence of, at least part of, the GR1 transcription unit had been determined.

Chapter 6. Chromosome mapping and DNA sequence analyses of GR1.

Section 6.1. Introduction.

Given the screening procedure used to isolate the cDNA clone pGR165 (Watson, 1986; section 3.2.7), it is important to determine if this cDNA clone represents a novel gene or a previously characterised gene, protein, or genetic locus. In addition to possibly saving considerable time and effort in analysis of the GR1 gene, any relationship with previously characterised genes, proteins or loci would help greatly in predicting the potential function(s) of GR1-encoded polypeptides during mouse development. Such predictions could then be tested experimentally.

Information concerning characterised genes and proteins is stored in the form of nucleic acid and amino acid sequence databases (for example EMBL and GENBANK). The genetic linkage map of the mouse (Davisson and Roderick, 1986) provides a source of information concerning characterised genetic loci.

Relationships can be established by computer-assisted comparison of the nucleic acid sequence, amino acid sequence and chromosomal location of GR1 with the respective data bases. However, since none of these data bases is in any complete, it was entirely possible that GR1 would be found to represent a novel gene. Even in this case, the information generated would be useful in the design of subsequent experiments to determine the function of GR1.

The genetic code ensures that the amino acid sequence of a polypeptide chain has an essentially linear relationship with

the nucleic acid sequence of the mRNA from which it was derived. This nucleotide sequence is represented in cDNA clones. Thus, determination of the nucleotide sequence of GR1 cDNA clones would allow access to both nucleotide and amino acid data bases (section 6.1.2). The chromosomal location of the GR1 locus could be determined by segregation analysis (section 6.1.3).

Section 6.1.1. Determination of nucleic acid sequences.

There are essentially two methods which can be used to determine the nucleotide sequence of a DNA molecule:- chemical cleavage (Maxam and Gilbert, 1977) or the enzymatic method of chain termination (Sanger et al., 1977). Developments in vector technology and improvements in chain termination reactions have made the latter the technique of choice for sequencing projects greater than a few hundred base pairs.
The chain termination method is based on the principle that the template-dependent activity of Klenow fragment of E.coli DNA polymerase I has an absolute requirement for a free 3'OH terminus at the site of elongation of the oligonucleotide chain. When incorporated into the chain, 2',3' dideoxynucleoside triphosphates inhibit polymerase activity, causing a termination of chain elongation (Sanger et al., 1977). A single stranded template DNA molecule is used in the reaction along with a specific oligonucleotide primer to

provide an initial 3'OH terminus for chain elongation. Hence, when a specific dideoxynucleoside triphosphate is included, along with all 4 deoxynucleoside triphosphates, one of which is radiolabelled, a variety of different sized elongation products are produced with a common 5' end but varying sites of termination. Denaturing polyacrylamide thin gel electrophoresis (Sanger and Coulson, 1978) of the products of termination reactions, on a given template, specific for each dideoxynucleotide allows the radiolabelled elongation products to be separated according to size. Thus, the order of nucleotides along the template DNA molecule can be determined (Sanger et al., 1977).

6.1.1.(i) Development of sequencing vectors.

Initially, the chain termination method made use of single-stranded template DNA molecules in which strand separation was achieved by physical methods (Sanger et al., 1977). However, such procedures are laborious and inefficient. The requirement for single-stranded DNA templates was met by the development of the filamentous single stranded bacteriophage, M13 (Pratt, 1969), as a cloning vector for DNA sequence determination (Messing et al., 1977; Messing, 1983). These vectors provide a biological method of strand separation.

On infection of a host E.coli cell, the single-stranded DNA within the M13 bacteriophage particle converts to double

stranded DNA, the replicative form (RF) of M13 (Messing, 1983). Both RF and single stranded M13 DNA can be purified, making M13 suitable for both DNA sequence determination and amenable to restriction enzyme technology.

Unlike bacteriophage lambda (section 4.1.1.), all genes in M13 are essential. However, insertional mutagenesis selection procedures led to the identification of a non-essential part of an intergenic region at which additional DNA sequences could be introduced into the M13 genome (Messing et al., 1977). Subsequent to the development of these first M13 vectors (Messing et al., 1977), a variety of modified vectors has been developed (Messing and Vieira, 1982; Norrander et al., 1983; Yanisch-Perron et al., 1985), all of which are designed to facilitate easy subcloning of DNA restriction fragments.

These M13 vectors contain a polylinker sequence, encoding numerous restriction enzyme recognition sequences, in frame with the E.coli lacZ gene. The absolute requirement of the F-pilus for M13 infection is met by the use of a host E.coli strain bearing an F'lac,Δpro plasmid. Selection for proline prototrophy on minimal agar maintains this F' plasmid. A deletion in the lac operon of the host E.coli strain facilitates the selection of recombinant M13 clones by an in situ colour reaction using X-gal.

In this assay, infected cells are grown in the presence of X-gal and the lac inducer IPTG. In non-recombinant bacteriophage, the M13 lacZ gene is induced to produce

functional β-galactosidase. The β-galactosidase cleaves X-gal to produce the deep blue dye dibromochloroindigo, giving a blue appearance to the M13 plaque. Recombinant M13 clones produce no β-galactosidase as a consequence of the restriction fragment to be sequenced being inserted into the polylinker sequence, disrupting the lacZ gene. Thus, recombinant M13 plaques appear colourless. The polylinker sequence also facilitates forced cloning (Messing, 1983), whilst recombinant clones can also be detected by plaque hybridisation techniques (Messing, 1983). The relative orientations of the inserted fragment in M13 can be established by comparison of the termination pattern of different clones with a given dideoxynucleoside triphosphate (section 2.22.2).

Whilst up to 7 times the unit length of the M13 genome can be packaged (Messing, 1983), DNA fragments >2kb are often found to be unstable (Sanger et al., 1980; Messing, 1983). Protocols have been developed which facilitate DNA sequence determination by the chain termination method using double stranded DNA in plasmid vectors (Korneluk et al., 1985; Yanisch-Perron et al., 1985). More recently, the chain termination method has been used to determine the DNA sequence of specific regions of the genome amplified by the method of the polymerase chain reaction (Mullis and Faloona, 1987), obviating the use of cloning vectors (Stoflet et al., 1988).

6.1.1.(ii) Modifications of the chain termination method.

Since the development of the chain termination method (Sanger et al., 1977), this technique has been improved in a number of ways. The use of buffer gradient gels (Biggin et al., 1983) and field gradient gels (Ansorge and Labeit, 1984; Olsson et al., 1984) has allowed more sequence to be determined from a single polyacrylamide gel. ^{32}P has largely been replaced by ^{35}S as the radiolabel of choice. The shorter path length of β-particles emitted by ^{35}S results in sharper definition of radiolabelled bands and a lower incident radiation dose is received by the worker. Also, radiolysis of ^{35}S radiolabelled DNA is markedly reduced, compared with ^{32}P, allowing reaction products to be stored up to one week without significant degradation (Biggin et al., 1983).

The formation of stable secondary structures in radiolabelled products of chain termination reactions often results in the appearance of compressions of radiolabelled bands following electrophoresis, from which the order of nucleotides cannot be determined. Additionally, secondary structures in the template DNA may result in non-specific sites of termination giving a similarly uninterpretable banding pattern. However, such regions are likely to occur at different positions in each of the two complementary strands in a DNA molecule (Sanger et al.,1977) and so it is important to determine the nucleotide sequence of both strands. Alternatively, replacement of dGTP with dITP (Mills and Kramer, 1979) or deoxy-7-deazaguanosine

(Mizusawa et al., 1986) in the chain termination reactions has been found to reduce secondary structures caused by G-C rich regions.

The original chain termination procedure used Klenow fragment of E.coli DNA polymerase I for primer extension. However, this enzyme has a low processivity (Kornberg, 1982), a 3′ to 5′ exonuclease activity and is known to discriminate 1000-fold against incorporating dideoxynucleotides (Atkinson et al., 1969). AMV reverse transcriptase has been used in the chain termination method (Chen and Seeburg, 1985). Whilst the properties of this enzyme (Kornberg, 1982) make it more suitable for chain termination reactions, it has a slow reaction rate.

Recently, a chemically modified bacteriophage T7 DNA polymerase (Tabor and Richardson, 1987), available commercially as Sequenase (section 2.22.4), has been developed. This enzyme is highly processive, shows no discrimination between nucleotides, has a fast reaction rate and the chemical modification has removed the exonuclease activity of T7 DNA polymerase (Tabor and Richardson, 1987).

Whilst the modifications discussed above, along with many others, have optimised the chain termination method, the limiting factor in this method is the number of nucleotides that can be determined for a given clone by denaturing polyacrylamide thin gel electrophoresis. Thus, large sequencing projects generally involve the assimilation of overlapping sequences for numerous M13 clones. Protocols have

been developed which enable unidirectional deletion of sequences in M13 RF DNA, facilitating the generation of M13 clones containing overlapping DNA sequences from a single initial clone (Yanisch-Perron et al., 1985). Protocols are also available which enable unidirectional deletions to be introduced into single stranded M13 template DNA (Dale et al., 1985).

However, such technology is not necessarily required. Given a detailed restriction enzyme map of the DNA molecule to be sequenced, it is possible to generate overlapping M13 clones simply by careful choice of the DNA restriction fragments to be subcloned. Generally, standard sequencing conditions can be used initially, some of the modifications given above only being used when specific problems are encountered.

With regard to GR1, the interest in determining the DNA sequence of this clone lies not only in gaining access to nucleotide data bases, but also in the analysis of any potential polypeptide-encoding capacity of this transcription unit. As mentioned earlier, this is most easily achieved by DNA sequence determination of GR1 cDNA clones. However, cDNA synthesis protocols make use of AMV reverse transcriptase for synthesis of cDNA from mRNA species (Watson and Jackson, 1985; Watson, 1986) and this is known to be prone to error.

The error rate of AMV reverse transcriptase has been calculated at 1/600 nucleotides on both synthetic (Battula and Loeb, 1975) and natural (Gopinathan et al., 1979) templates when dideoxynucleoside triphosphates are supplied in the range

of 25-50uM. Although significantly higher nucleotide concentrations were used in the construction of the cDNA library from which the GR1 cDNA clone was isolated (Watson, 1986) it has been estimated that, under conditions used for cDNA synthesis, 50% of the cDNA clones of a 2kb mRNA species of moderate abundance will contain a single base error (Gerard, 1986). Any such errors in the nucleotide sequence of open reading frames in GR1 cDNA clones could radically alter the predicted amino acid sequence of putative GR1-encoded polypeptides, making analysis futile. Thus, if there are any doubts concerning the potential amino acid sequence encoded by cDNA clones, the DNA sequence of the corresponding region of the genomic locus should be determined.

Section 6.1.2. Analysis of nucleic acid and amino acid sequences.

Given that the sense strand of the GR1 cDNA clone pGR165 is known (section 5.2.2), once the nucleotide sequence of coding regions of the GR1 transcription unit had been determined, it would be possible to predict the amino acid sequence of putative polypeptides encoded by GR1. Both nucleic acid and amino acid sequences could then be compared with data bases and subjected to more detailed analyses.

6.1.2.(i) Searching data bases.

Both nucleic acid and amino acid sequence data bases have been established (for example EMBL and GENBANK) in which all published sequences are stored. These are available commercially and are periodically updated. Amino acid data bases consist of amino acid sequences derived directly from purified peptides, as well as hypothetical polypeptides derived by conceptual translation of nucleic acid sequences. Whilst the data bases at EMBL and Los Alamos were established separately, there is now a concerted effort to conform the information contained in data banks at both locations. As such, there are now few sequences unique to one source. At the time of writing the data bases available are the October 1987 and July 1987 releases of EMBL nucleotide and amino acid data bases respectively.

Although current data bases are large (14397 nucleic acid and 5205 amino acid sequences in the current EMBL data bases), they contain much redundant data in the form of many variants of closely related genes and proteins. In addition the data bases are biased, there being a disproportionate amount of data from some species relative to others. Nevertheless, the first step in analysis of DNA or amino acid sequences is to compare the newly derived sequence with those contained in the data bases. Given the large amount of data involved, computer software has been developed for such analyses.

Essentially, these programs compare a string of characters,

the length of which can be chosen as a "window", of the target sequence with all strings of characters, of the same window size, in the data base. The window then moves one nucleotide further along the target sequence and the comparisons are repeated. Only when matches are observed in a series of windows is a match recorded.

The data base search programs are designed to detect homologies between nucleic acid or amino acid sequences and so the comparisons take into consideration substitutions, insertions, deletions and duplications of sequences which may have occurred across evolutionary time. Such changes are accommodated by allowing gaps between homologous sequences. However, since the programs are designed to find maximum homologies, gapping could get out of control. This is prevented by the use of a gap-penalty which reduces the overall score attributed to a match between two sequences. Most programs order the matches such that only the highest values are shown as "hits".

The large size of data bases results in such comparisons being very slow. However, data base searching can be speeded considerably by the use of an algorithm based on matching k-tuples (Wilbur and Lipman, 1983). This sets a value, k, for the minimum number of matches detected within the window for a given sequence. Only if this value is reached are comparisons around this window investigated.

Since only 4 characters are involved, searching nucleic acid data bases has a low signal:noise ratio. With no gap

considerations, 25% homology between two sequences would be expected by chance. Thus, many short alignments, such as between AT rich nucleic acid sequences, would be expected and are generally disregarded. The use of 21 characters (20 amino acids and a termination) to define amino acid sequences results in a much higher signal:noise ratio when searching amino acid data bases.

Should an overall homology, or identity, be detected for a newly derived sequence, this information will be extremely useful in the choice and design of further experiments concerned with analysis of that gene. However, even if the sequence shows no global homologies, further analysis can produce informative data.

6.1.2.(ii) Nucleic acid consensus sequences.

Comparative analyses of eukaryotic nucleic acid sequences have established the presence of particular consensus sequences thought to be involved in certain aspects of DNA metabolism. Elements such as TATA boxes, GC rich elements, and protein binding domains, are found in the 5′ regulatory regions of genes (section 1.5). Consensus sequences have also been found at exon/intron junctions and a polyadenylation signal, AATAAA, has also been characterised (Breathnach and Chambon, 1981). Other consensus sequences, such as ATTTA, are common in the 3′ untranslated region of many cytokine and proto-oncogene mRNAs

and may confer RNA instability in certain conditions (Cosman, 1987). In the context of the current work on GR1, however, it is analysis of the amino acid sequence of putative GR1-encoded polypeptides which is of major interest in attempting to establish potential functions for GR1.

6.1.2.(iii) Protein structure predictions.

A central tenet of molecular biology is that protein structure is related to function. However, despite a large amount of effort, it is not yet possible to predict, with any degree of certainty, the tertiary structure of a polypeptide solely from amino acid sequence data. Attempts to predict the tertiary structure of proteins from amino acid sequences involve either the use of simulated folding models or empirical methods by inference from known protein structures (Taylor, 1986). Both of these approaches have severe limitations.

Simulated folding models rely on the the use of universal folding laws from which protein structures can be derived by extrapolation. Such laws are concerned with the minimum energy state of the protein. However, it is not yet possible to predict the most stable conformation of all domains of a protein since the conformation of a domain may be dependent on those of all others. The limited capacity of current computer technology makes simulations slow, and the role of water,

believed to be a major force in protein folding, cannot yet be considered (Taylor, 1986). None of the model building approaches based on energy calculations have been able to predict a structure close to that of the native protein (Cohen and Sternberg, 1980).

The empirical approach can produce reliable predictions. However, it is limited since structure predictions can only be made from homologies with known protein structures. The derivation of such structures relies solely on X-ray diffraction patterns of protein crystals. The limiting step in this process is that of crystal formation and, as such, only 200 or so, mainly globular, proteins have been analysed in this way (Taylor, 1986). Such analyses have revealed a structural hierarchy of subunits in proteins (Richardson, 1981).

Secondary structure is the result of steric restraints of the peptide bond giving two stable configurations:- the α-helix and the β-sheet (Pauling et al., 1951). Super-secondary structure (Rao and Rossman, 1973) consists of sequentially adjacent regions of secondary structures which are packed together in an energetically favourable conformation. Such conformations include the $\beta\alpha\beta$ unit (Rao and Rossman, 1973; Sternberg and Thornton, 1976), the β-hairpin and β-meander (Taylor, 1986) and the 4-α-helical bundle (Weber and Salemme, 1980). Protein domains are structurally or functionally intact structures, larger than secondary structures, and include all-α, all-β and α/β alternations (Richardson, 1981).

Computer programs have been developed (Chou and Fasman, 1974; Garnier et al., 1978) which compare the newly derived amino acid sequence with that of known folding units, giving predictions of regions of α-helix, β-sheet, turns and coils. The Robson method (Garnier et al., 1978) has the advantage over that of Chou and Fasman (1974) in that the former predicts the secondary structure of a given residue by considering the residue in the context of a "window" including 8 residues either side. Decision constants (DC) are applied to bias a predicted structure further towards one type, if one type of structure is found to predominate in a sequence. Whilst DC values are seen to work well for all-α and all-β domains, α/β domains are problematical. The refinements of the Robson method introduced by Taylor and Thornton (1983, 1984) significantly improved the accuracy of prediction of super secondary structure of α/β domains.

Properties of the side chains of amino acids also play a vital role in protein folding. Protein structures generally consist of hydrophobic domains within a hydrophobic shell. Programs are available to predict patterns of hydrophobicity along an amino acid sequence (Kyte and Doolittle, 1982). Local patterns of hydrophobicity can be used to predict secondary structures (Lim, 1974) and can be usefully combined with the statistical methods discussed above (Cohen et al., 1983).

It should be remembered that the analyses described have been shown to be only 50-55% accurate when predictions are compared with crystallographic data of known proteins (Kabsch and

Sander, 1983; Nishikawa, 1983), although the modifications of Taylor and Thornton increased this accuracy significantly (Taylor and Thornton, 1983, 1984).

Whilst secondary structure predictions should always be interpreted with caution, they can facilitate a functional classification of the protein under investigation. Statistical analysis of known protein structures has shown a strong correlation between the predominant folding pattern and function (Nishikawa and Ooi, 1982; Nishikawa et al., 1983). Thus, intracellular enzymes are rich in α/β structures, intracellular non-enzymes in α-helical structures. Extracellular enzymes and non-enzymes are rich in $\alpha+\beta$ and β structures, respectively (Nishikawa et al., 1983; Taylor, 1986).

As with nucleic acid sequences, there are consensus amino acid sequences such as those of active sites and binding domains (Doolittle, 1986) which may be useful in designing subsequent experiments concerned with functions of the novel gene.

6.1.2.(iv) Prediction of antigenic determinants.

When using the protein structure predictions discussed above, it is easy to forget that, in the case of a gene such as GR1, analysis is of a hypothetical amino acid sequence. It is important to confirm experimentally that this predicted amino acid sequence is the true translation product of GR1 mRNA.

Whilst definitive proof of identity between the predicted and true amino acid sequence of GR1 would require purification and direct sequence determination of the GR1 protein, an easier route is offered by the production of antisera against peptides of the predicted GR1 polypeptide. In addition, this approach may be more fruitful than that of fusion proteins to produce GR1 specific antisera (section 5.3.1). Several rules have been established concerning the choice of synthetic peptide likely to correspond to antigenic sites on the native protein (Walter, 1986).

Synthetic peptides corresponding to predicted amino terminal and carboxy terminal peptides often serve as antigenic sites (Walter et al., 1980) since these regions are frequently exposed on the native protein and generally contribute little to overall folding of the polypeptide chain. In contrast, internal domains are likely to be fixed within a three dimensional structure and so may not be as flexible as synthetic peptides of such regions (Walter et al., 1980; Walter, 1986).

Predictions of likely antigenic sites in amino acid sequences can also be based on the hydrophilic properties of polypeptide domains (Hopp and Woods, 1981) or on predictions of the segmental mobility of a peptide sequence (Tainer et al., 1984; Westhof et al., 1984). However, since the same peptide sequence may adopt different conformations in the context of different polypeptide chains (Kabsch and Sander, 1984) synthetic peptides may not always give rise to useful

antisera.

Thus, whilst it may not be possible to establish any overall homology between GR1 and other genes or proteins, determination of the nucleotide sequence of coding regions of this transcription unit can be of great use in subsequent analyses of gene function.

Initial DNA sequence data of the GR1 cDNA clone pGR165 established that this was a novel gene. Subsequently, in parallel with expression studies (chapters 3 and 5), and attempts to raise antisera to GR1 fusion proteins (section 5.3.1), the sequence of 2674 nucleotides of the GR1 transcription unit was determined. This sequence encodes 247 amino acids of the putative carboxy terminal domain of the GR1 polypeptide.

Section 6.1.3. Linkage map of the mouse.

Determination of the chromosomal location of a given genetic locus is dependent upon the use of inbred strains of mice. An inbred strain is defined as one in which sister/brother matings have been carried out for at least 20 generations and it can be considered as being homozygous for all loci, with the exception of spontaneous mutations.

The origin of inbred strains of the laboratory mouse can be traced back to the efforts of a small number of researchers in the early twentieth century. Most of the 200 or more inbred

strains now available (Green, 1981) are believed to be derived from a small number of founder animals (reviewed by Hogan et al., 1986).

In order to map a given locus, there must be two or more alleles which are segregated between different inbred strains. Provided the alleles can be distinguished, it is possible to determine linkage between 2 loci by analysis of the inheritance of the alleles on back cross analyses of inbred strains homozygous for each of the alleles. Distinction of the alleles of a locus between different inbred strains can be made according to phenotype, or by biochemical or DNA polymorphisms.

Many alleles can be distinguished on the basis of mutations they induce in the mouse. Some mutations, such as coat colour, hair morphology or pigment patterns, are derived directly from the early strains developed by mouse "fanciers". Others were isolated as recessive mutations during early inbreeding experiments. More recent work has established a variety of radiation and chemically induced mouse mutants (Green, 1981). A wide variety of abnormalities are available, some of which can be traced to abnormal embryogenesis. Of particular relevance in the present study, given the expression pattern of GR1 (section 5.2.4), is the work of Hans Gruneberg who investigated the phenotypes of many mutant strains in which skeleton development was defective (Gruneberg, 1963; Johnson, 1986). Of more than 700 loci described in the mouse, 95 are known to have phenotypic effects on the skeleton (Green,

1981).

Initially, mapping loci was limited to analysis of the segregation of alleles in back cross analyses between 2 different inbred strains. Distinction of the alleles can be made on the basis of morphology, or biochemical variants of proteins, or DNA restriction fragment length polymorphisms between the two progenitor strains. However, such analyses are expensive, time consuming and the data derived from them are not cumulative. A number of additional approaches are now available which together have facilitated the construction of an increasingly detailed linkage map of the mouse genome.

The establishment of recombinant inbred (RI) strains of mice (Bailey, 1971; Taylor, 1981) has greatly aided linkage analysis. Recombinant inbred strains are derived by systematic inbreeding of the progeny of a cross between two pre-existing progenitor inbred strains. On the assumption that there is no selection, a genetic locus will become fixed for one or other of the alleles at that locus.

Whilst RI strains can be used to analyse gene action (Swank and Bailey, 1973), they can also be used for gene mapping (Taylor, 1975). The major advantage of RI strains for gene mapping studies is that the data are cumulative. Comparison of the strain distribution pattern (SDP) of the alleles of a given locus can be compared with that of all other loci for which the SDP has been determined. Loci which are linked will tend to co-segregate in the RI strains. The distribution of loci in RI strains is the result of an average of 4 meioses

(Taylor, 1978). This is in contrast to the single meiotic event analysed in a conventional backcross. The practical consequences of this lie in the finding that only relatively tight linkage relationships, <12cM, can be established using RI strains (Taylor, 1978). Thus, whilst RI strains are extremely useful for detailed mapping studies, linkage between loci >12cM apart must be established by other means.

There are other limitations to the use of RI strains for linkage analyses. Some of the RI strains are very poor breeders and have died out, resulting in some series of RI strains being incomplete. Also, only loci at which progenitor strains differ can be analysed. This problem is alleviated, to a degree, by the availability of RI strains derived from different combinations of progenitor strains (Taylor, 1981), but the common origin of inbred strains still limits the polymorphic variations in RI strains.

This limitation has been resolved, to a large extent, by the use of interspecies backcrosses between inbred strains of Mus musculus domesticus and Mus spretus (Robert et al., 1985). In addition to increasing the probability of finding polymorphisms at the locus of interest, non-random segregation of loci has not been detected in these interspecific hybrids. In RI strains, there may be selection for or against certain arrangements of genes. Since backcross analysis is involved, linkage of loci up to 30cM apart can be established (Robert et al., 1985).

DNA probes can also be used to map loci by hybridisation

analyses of interspecific somatic cell hybrids, such as Chinese Hamster x mouse cell hybrids (Lalley et al., 1978). Tritiated DNA probes have been used to assign a locus to a specific region of a chromosome by means of in situ hybridisation to metaphase chromosome preparations (Gerhard et al., 1981).

By a combination of all these various techniques it has been possible to construct a fairly detailed linkage map of the mouse genome (Davisson and Roderick, 1986).

Comparative analyses of the autosomal linkage maps of mouse and human have shown that the 184 autosomal loci mapped to date in both species fall into 40 conserved segments (Searle et al., 1987). The majority of these conserved syntenic groups are small, average size being approximately 8cM (Nadeau and Taylor, 1984), suggesting that they represent the relics of ancient linkage groups not yet subjected to recombination. However, some of these segments are in the order of 20-25cM in size (Searle et al., 1987), indicating that their persistence may be greater than that expected by chance. Some chromosomes have been found to be highly rearranged, such as human chromosome 7 which contains conserved segments from at least 6 mouse chromosomes. In some cases, loci appear to have intercalated within a syntenic group of loci (Searle et al., 1987).

It is also interesting to note that, in some cases, related genes are often clustered within a specific region of a chromosome. Examples include the Hox-1 and Hox-2 loci of the

mouse (Martin et al., 1987) and the genes encoding mouse salivary proline rich proteins (Ann and Carlson, 1985) which map close to the mouse taste loci (Azen et al., 1986). Several growth factors, including IL-3, GM-CSF and M-CSF, along with growth factor receptor genes for M-CSF and PDGF, are clustered to a region on the long arm of human chromosome 5 (Le Beau et al., 1986, 1987; Pettenati et al., 1987).

It should be emphasised that whilst over 700 genetic loci have been characterised in the mouse (Green, 1981), these represent only a very small proportion of the $0.5\text{-}1.0\times10^5$ genes estimated to be present in the mouse genome (McKusick and Ruddle, 1977). Moreover, linkage estimates, particularly those from RI analysis, usually have large confidence intervals (Silver, 1985). Thus, precise mapping of a locus is only possible when RI analysis is used in combination with other techniques. Only in a very few cases has it been possible to correlate mutations in characterised genes to known developmental mutant mouse strains (Dautigny et al., 1986; Molineaux et al., 1986; Readhead et al., 1987).

Thus, the mouse linkage map is in no way as useful a source of information as nucleic acid and amino acid data bases in attempts to establish relationships between a newly characterised gene and other genes, proteins or loci. However, given the developmental expression of GR1, as revealed by RNA in situ hybridisation (section 5.2.4), and the large number of skeletal abnormalities mapped in the mouse (Gruneberg, 1963; Johnson, 1986), determination of the chromosomal location of

GR1 was considered useful in attempts to establish potential functions for this novel gene.

For this purpose, DNA restriction fragment length polymorphisms at the GR1 locus between DBA2/J and C57bl/6 mice were established. These were then used to determine the strain distribution pattern of GR1 in BxD recombinant inbred strains (Taylor, 1981).

Section 6.2. Results.

Section 6.2.1. GR1 cDNA clones available for DNA sequence determination.

As discussed earlier (section 6.1.1), by far the easiest way to determine the potential amino acid sequence of the GR1 transcription unit is from analysis of open reading frames present in the sense strand of GR1 cDNA clones.

However, the cDNA clone pGR165, being 2.5kb in size, represents only a portion of the nucleotide sequence present in the 6.8kb GR1 mRNA (section 3.2.7). Whilst the size of untranslated regions within this mRNA could not be estimated, it was possible that pGR165 contained little or no amino acid encoding capacity.

Given the problems encountered in cloning the GR1 locus (section 4.3.2), genomic clones corresponding to the cDNA clone pGR165 were not available at the initiation of sequence determination. Thus, attempts were made to isolate cDNA clones containing sequences transcriptionally upstream of those found in pGR165. The K880 restriction fragment of pGR165 (fig. 4.2) was used as a radiolabelled probe to screen three murine cDNA libraries prepared as EcoRI-linkered cDNA molecules in the bactriophage lambda vector gt10 (Watson and Jackson, 1985).

Screening of 10^4 clones of the PCC3 cDNA library from which the original GR1 cDNA clone had been isolated (Watson, 1986), failed to isolate any clones containing sequences additional

to those present in pGR165 (data not shown). $7x10^3$ clones of a cDNA library prepared from RNA isolated from ES cells differentiated *in vitro* for 7 days (F. Poirier, pers. comm.) yielded two K880 homologous clones, neither of which contained cDNA inserts large enough to extend beyond the 5′ end of pGR165 (data not shown). Screening of $7x10^3$ clones of a cDNA library prepared from RNA isolated from ES cells differentiated *in vitro* for 10 days (F. Poirier, pers. comm.) led to the isolation of one K880 homologous clone, 10-34, which contained sequences not present in pGR165 (data not shown).

The 1.4kb cDNA insert of this clone contained an internal EcoRI restriction site. The 400bp EcoRI restriction fragment of 10-34 was found to have homology to both K880 and RK5 restriction fragments of pGR165 (fig. 4.2; data not shown). In contrast, the 1kb EcoRI fragment of 10-34 had no homology to any region of pGR165 (data not shown). However, Northern blot analysis revealed that the novel restriction fragment of 10-34 hybridised to a single mRNA species of approximately 2kb in both PCC3 cells and differentiated derivatives of this cell line (data not shown). No hybridisation to the characteristic 6.8kb GR1 mRNA seen in EC and ES cells (section 3.2.7; section 3.2.9) was observed.

On the basis of these data, it was concluded that it was likely that the cDNA clone 10-34 represented a double ligation event of 2 otherwise unrelated cDNA molecules, only one of which corresponded to a region of pGR165. Subsequently, it was

found that this novel fragment had no homology with the GR1 genomic clone N28 (data not shown).

Thus, despite screening $2.4x10^4$ cDNA clones from three libraries, no cDNA clones containing sequences upstream of those represented by pGR165 were isolated.

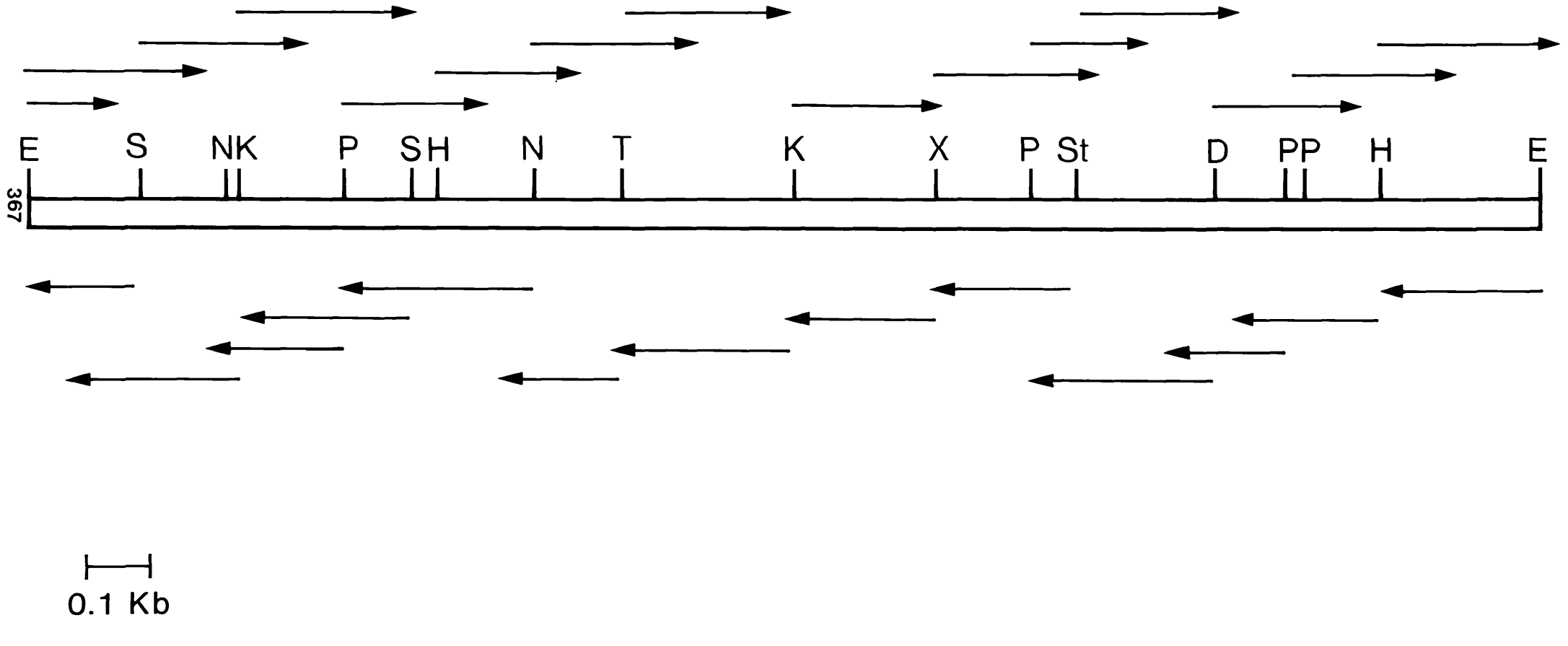
Section 6.2.2. Sequencing strategy for pGR165.

Given that pGR165 was the largest available cDNA clone of the GR1 transcription unit (section 6.2.1), a strategy was designed to determine the entire nucleotide sequence of this clone. As mentioned earlier (section 6.1.1), accurate prediction of the potential amino acid sequence encoded by a DNA sequence requires precise determination of nucleotide sequence. Single base insertions, deletions or substitutions, whether introduced during cDNA synthesis and sequencing reactions or by poor resolution of gels, could radically alter the predicted amino acid sequence, making analysis futile.

The sequencing strategy for pGR165 was designed to ensure that the entire nucleotide sequence of each strand of the cDNA clone was determined (fig. 6.1). Any ambiguities, caused by compaction of bases or non-specific terminations (section 6.1.1), on one strand would be unlikely to be encountered in the same region of the complementary strand. Additionally, the sequence derived for one strand could be used to confirm that

Figure 6.1. Sequencing strategy for cDNA clone pGR165.

Various restriction enzyme fragments of pGR165 were subcloned into M13 vectors mp8, mp9 or mp19, as described in section 2.16, and sequenced by the chain termination method (section 2.22). Arrows indicate the orientation and region of pGR165 determined for a given M13 clone. The complete sequence of each strand of the cDNA clone was determined by assimilation of data from overlapping clones. D= DraI, E= EcoRI, H= HincII, K= KpnI, N= NdeI, P= PstI, S= SacI, St= StuI, T= TaqI, X= XhoII.


E
S
NK
P
SH
N
T
K
X
P
St
D
PP
H
E
367
0.1 Kb

derived for the other, so increasing confidence that the nucleotide sequence had been accurately determined.

Under the electrophoretic conditions used (section 2.15), only 250-300 nucleotides from the primer binding site could be determined for a given M13 clone. Thus, a series of overlapping M13 clones of pGR165 was constructed. This demanded a restriction map for pGR165 considerably more detailed than that previously available (fig. 3.9). Initial sequence determination was carried out by subcloning restriction fragments into M13 on the basis of the original restriction map of pGR165 (fig. 3.9). Additional restriction sites predicted from these sequence data were then used to generate further M13 subclones. This not only confirmed predicted restriction enzyme sites, but made it possible to derive overlapping sequences of each strand of pGR165 as outlined in fig. 6.1.

The complete sequence of each strand of the cDNA clone was determined by the chain termination method using Klenow fragment of E.coli DNA polymerase I (section 2.22.3). Determination of the cytosine rich region between positions 410 and 493 (fig. 6.2) was based solely on sequence data of the lower strand. Non-specific terminations made determination of the sequence of the sense strand of this region impossible. Analysis of the determined sequence indicated that the sense strand of pGR165 (section 5.2.1) contained a single open reading frame from position 6-515, the synthetic EcoRI restriction site being excluded (fig. 6.2). However, two

Figure 6.2. Nucleotide sequence of cDNA clone pGR165.

The complete nucleotide sequence of the sense strand of pGR165 (section 5.2.1) is shown. Numbering begins at the first nulceotide of the synthetic EcoRI restriction enzyme recognition sequence at the 5′ end of the cDNA clone. Location of restriction enzyme recognition sequences for EcoRI, KpnI, PstI, SacI and StuI are shown. The amino acid sequences encoded by each of the three large open reading frames in this sense strand, between nucleotides 6-515, 263-649 and 658-951, are shown.

EcoRI 20 40
GAAT TCC CAG TGG ACA TTT GTA CTC AAT GTC TGA GTC TGA AAT GAA TGC TCT GCG

Pro Ser Gly His Leu Tyr Ser MET Ser Glu Ser Glu MET Asn Ala Leu Arg

60 80 100
AAA TTT CGT GGA CAG GAA TGT TAA AGA TGG GCT CAT GAC TCC CAC TGT GGC GCC

Asn Phe Val Asp Arg Asn Val Lys Asp Gly Leu MET Thr Pro Thr Val Ala Pro

120 140 160
CAA TGG AGC CCA AGT CCT GCA AGT GAA AAG AGG GTG GAA ACT CCA AGT CAC TTA

Asn Gly Ala Gln Val Leu Gln Val Lys Arg Gly Trp Lys Leu Gln Val Thr Tyr

SacI 180 200
CAA TTG CCG AGC TCC ACA GAG TGG CAC CAT CCA AAA TCA GTA CCT ACG CAT GTC

Asn Cys Arg Ala Pro Gln Ser Gly Thr Ile Gln Asn Gln Tyr Leu Arg MET Ser

220 240 260
TCT TCC AAA TAT TGG AGA CCC TGC ACA CCT GGC AAG CTA TGG TGA ATT TGT CCA
Ile Cys Pro
Leu Pro Asn Ile Gly Asp Pro Ala His Leu Ala Ser Tyr Gly Glu Phe Val Gln

280 300 320
AGT TCC TGG CTA CCC ATA TCC AGC CTA TGT TTA CTA TAC AAG CCC GCA TAT GAT
Ser Ser Trp Leu Pro Ile Ser Ser Leu Cys Leu Leu Tyr Lys Pro Ala Tyr Asp
Val Pro Gly Tyr Pro Tyr Pro Ala Tyr Val Tyr Tyr Thr Ser Pro His MET MET

KpnI 340 360 380
GAC TGC GTG GTA CCC AGT AGG ACG AGA TGT ACA TGG ACG AAT AAT CGT TGT GCC
Asp Cys Val Val Pro Ser Arg Thr Arg Cys Thr Trp Thr Asn Asn Arg Cys Ala
Thr Ala Trp Tyr Pro Val Gly Arg Asp Val His Gly Arg Ile Ile Val Val Pro

400 420
TGT TGT AAT CAC CTG GTC TCA AAA TAC GAA CCG CCA GCC TCC GGT GCC CCA GTA
Cys Cys Asn His Leu Val Ser Lys Tyr Glu Pro Pro Ala Ser Gly Ala Pro Val
Val Val Ile Thr Trp Ser Gln Asn Thr Asn Arg Gln Pro Pro Val Pro Gln Tyr

440 460 480
TCC TCC TCC GCA GCC ACC TCC ACC ACC ACC ACC ACC TCC ACC GCC ACC ACC ACC
Ser Ser Ser Ala Ala Thr Ser Thr Thr Thr Thr Thr Ser Thr Ala Thr Thr Thr
Pro Pro Pro Gln Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro

```
              500    PstI                520                         540
TCC ACC AGC ATC ATC CTG CAG TGC TGC GTA GAA CCT GTC ATG TCC TTT GTA GTC
Ser Thr Ser Ile Ile Leu Gln Cys Cys Val Glu Pro Val MET Ser Phe Val Val
 Pro Pro Ala Ser Ser Cys Ser Ala Ala

                       560                         580
TCT GCC CTC AAC TTG ATC CTG TGC AGC TTC TCA ATC TAT GAC TGT GTG GTA CTG
Ser Ala Leu Asn Leu Ile Leu Cys Ser Phe Ser Ile Tyr Asp Cys Val Val Leu

   600                   SacI  620                         640
GAC CTT CAG AGG CGC ACA GAG CTC AAG TCA GTT TTC GTC TTG ACT GCC ACT TTA
Asp Leu Gln Arg Arg Thr Glu Leu Lys Ser Val Phe Val Leu Thr Ala Thr Leu

           660 StuI                    680                         700
TAA GTT GAC AGG CCT GGG TTT TAC TTG TTA AAA CCT CTC ACC ATC TCA ATC ACA

          Gln Ala Trp Val Leu Leu Val Lys Thr Ser His His Leu Asn His Arg

                    720                         740
GGC TGC CAA GTG TCT TTA CAA AGA AGC TGA TAC AAA CAC AGG CCA TGC TGA TTT

  Leu Pro Ser Val Phe Thr Lys Lys Leu Ile Gln Thr Gln Ala MET Leu Ile Ser

760                         780                         800
CTT ACA GAG GGA GAG AAG AGG AAG AGA AGA AGA AAG AGG AGG AAG AGG ACA TGA

  Tyr Arg Gly Arg Glu Glu Glu Glu Lys Lys Lys Glu Glu Glu Glu Asp MET Thr

       820                         840                         860
CTT GCC CAT ATG CTG GGC ACC TTA TAA AGG AAG CCA GAC TTT TCG GTG CAG TAT

  Cys Pro Tyr Ala Gly His Leu Ile Lys Glu Ala Arg Leu Phe Gly Ala Val Trp

                880                         900                         920
GGA AAG GCT TCC GTG ATT CTC TTG CTG CAC CCC ACG AAA CTT CAC CAC CTT CAA

  Lys Gly Phe Arg Asp Ser Leu Ala Ala Pro His Glu Thr Ser Pro Pro Ser Asn
```

940 960
ACT CCA TTT TCA CGG TTC CGT TAA TTT TCA AGG AGC AGC AAC TCG ACT GGT TCT

Ser Ile Phe Thr Val Pro Leu Ile Phe Lys Glu Gln Gln Leu Asp Trp Phe Ser

980 1000 1020 1040
CTG CTA CATGAAACACCTCAGCTTGAAAAGGAAGTGCTCTCTCAGACTGACTTGTGAGTGTGCCTTCAC

Ala Thr

1060 1080 1100
ATTCTGGTGCAAATCATGTGTACCCAAGAACTCTGACATAGCATCTTACCATCATCATGCCAGGAATTGGT

1120 1140 1160 1180
TTCCTTAAAGTGTGACACTTGGCAACCAGAAGCGTCTGGTGGCTTCAGGTGTGTTAGGGTTTTGTAGAAAG

1200 1220 1240 KpnI
TTGTAACCACCTCTCAGTCATGCAAGAGGGTACATCAAAAAGACTTCTAAATTTGGTACCCAGACTTTGAC

1260 1280 1300 1320
AAGAGCCTGTGAGTTAGTTCATTGTAACCAACTTGTCAGGAACTTCAGCAAGAATTTTCATGAATGTTGTT

1340 1360 1380
AATTGGCTGGAGAAATACCAGCAAAGTTTTGCATCATCCACCCATTTGTCTGCCATTGTGCATCTGGACAT

1400 1420 1440 1460
CCATCAACCGCCCACTCAGTGGCATCTACCATGAAGATGTTTAGAGACGGGAAATGATTTTGCACCCAGAT

1480 1500 1520 1540
CCACACTCCTCTCATGTCAGATATTTGGAATTAGAGCACCGGGAATCGCCTTGAAGATATGAACTTCTAGA

1560 1580 1600
AATGGAGTCTCTTGCTACTACCTTGAAACATTTGTTTATCCTCTCTTTCTGTTTGATTCTACTACTACCAT

1620 PstI 1640 1660 1680
TAACCCTGCTGCAGGATTTGTCACCATTCTGCCTGACTGCTGAGACTCCATTTTGCTGCTATGCAAGGAGA

StuI1700 1720 1740
TGAAAGAGGCAGGCCTAAAGGAGCAGTAGGACATAACATGGTTTTTATTTCTATCTGTCATGATCTTAATG

1760 1780 1800 1820
GTGAAGTTTCCTTTTTCGTCAGCATTTATCTCTTTGTTATCCTGACATGTTTTAATTAGTTTAGTGGGTTT

1840 1860 1880
TTTTTTTTCTATTGGTGGTGGTGTTCTTTTGTTGTTGTTTTTTGTGGTCGTCATTTTGATTTTGGATCACT

1900 1920 1940 1960
TCGTGTTTTACAGTAATTACTTTTAAATGGTGCATTTGCTTCTGATTTTTTTTTTTTTATGAAGCATCACA

1980 2000 2020 PstI
TCAGTTTACCTCATATCTCAATTCCATCCTTCATGCATTTTTTTTTTAACTCATTTGATCTTCAAGCTGCA

2040 2060 PstI 2080 2100
GAGGGCCTAGCAATGGGTCATCACCTGCAGCCCTGGCATGTACACACGGACATTTGCCACCACTGAAAGCA

2120 2140 2160 2180
AACAGTTGGAGAAGTTGGCCACCTGAGTCAAGGAGGTGGTGCTGGTGTGAGTTCACAGCTCACAGGGGACG

2200 2220 2240
GTGGAACGTTGACATTGACTTTGGCAGTGTGTACTATGCTCTACTCCTATATATACTCTATAGATGTTAGG

2260 2280 2300 2320
CATTAAGGATAAGTGATCTTAAATTTACTGAAATTTTGTTAAGTTGATTAGATTTACATTAATTGTTTATA

2340 2360 2380
TTGTGTTTTTTCTTTTATTTTAGCACTTCCAACAAAAATGCCTTTATTCATTTATTAGGAAGAAATTGGAG

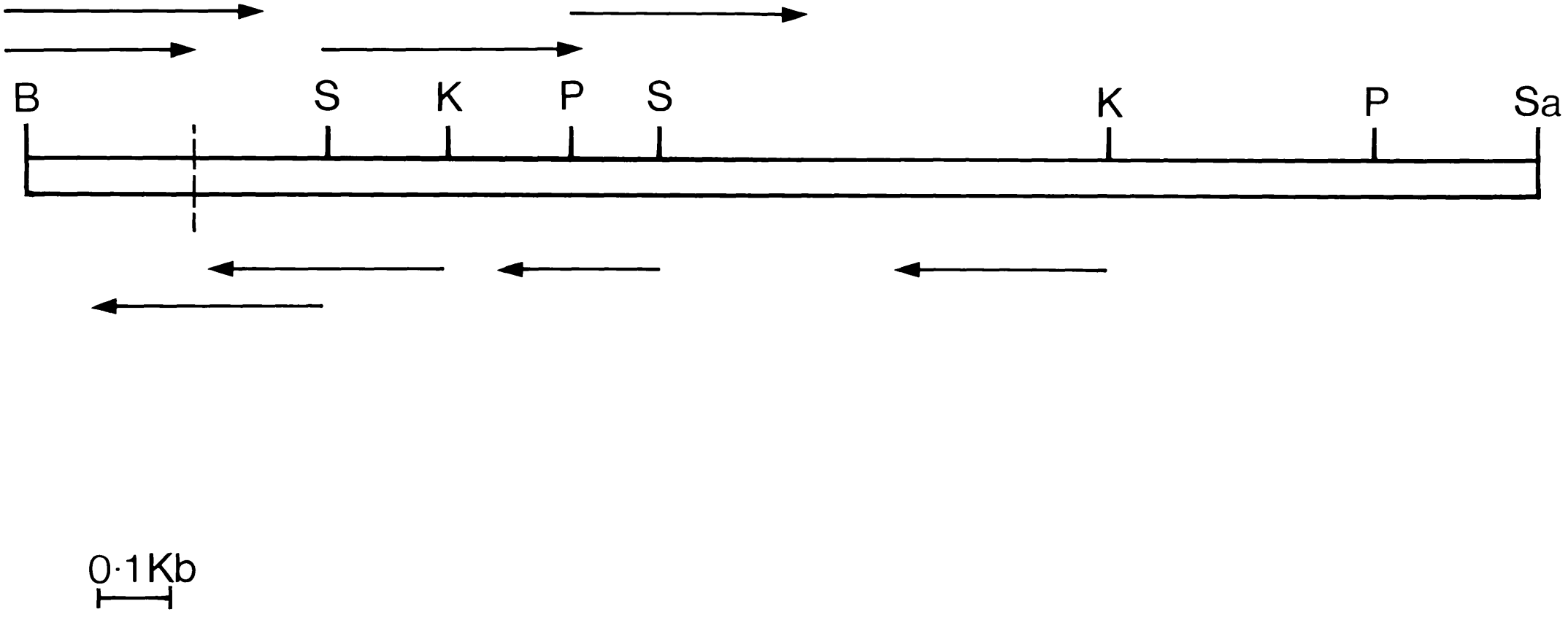
2400 2420 2440 EcoRI
TGGCGAACACAAACTAGCAAAATTAATTAATTGGCTGTGGGCCCAAATTTGAATTC

additional large open reading frames, from nucleotides 263-649, and 658-981, were present in this sense strand (fig. 6.2). Downstream of position 981 only small open reading frames were found. Given that single base insertions, deletions or substitutions could alter the predicted open reading frame, further sequence determination was carried out. The M13 clones used to determine the sequence of both strands of pGR165 upstream of the KpnI restriction site at position 1244 (fig.6.1 and 6.2), were sequenced again by the chain termination method but now using the modified T7 DNA polymerase, Sequenase (section 2.22.4). As discussed earlier, this enzyme is optimised for DNA sequence determination by the chain termination method (section 6.1.1). Although this method allowed determination of the cytosine rich region on both strands of pGR165, the DNA sequence determined was identical to that derived previously, confirming the three large open reading frames in the sense strand of pGR165 (fig. 6.2).

Section 6.2.3. Determination of partial DNA sequence of GR1 genomic clone N28.

Whilst the cDNA sequence determined for pGR165 (fig. 6.2) was considered accurate, it remained possible that copying errors by AMV reverse transcriptase during cDNA synthesis (section 6.1.1) had resulted in the DNA sequence of pGR165 not being identical with that of the corresponding region of the GR1

Figure 6.3. Sequencing strategy for genomic clone pBS28-A. Various restriction fragments of the N28 subclone pBS28-A (section 6.2.4), subcloned into M13 vectors mp18 or mp19 as described in section 2.16, were sequenced by the chain termination method using Sequenase (section 2.22.4). Arrows indicate the orientation and region of pBS28-A determined for a given M13 clone. The broken line indicates the position of the 5′ EcoRI linker in the cDNA clone pGR165 (fig. 6.1). B= BamHI, K= KpnI, P= PstI, S= SacI, Sa= SalI.


B
S
K
P
S
K
P
Sa
0·1Kb

transcription unit. Such potential errors seemed particularly relevant given the presence of 3 large open reading frames in the 5′ region of this cDNA clone (fig. 6.2). Thus, a strategy was designed to make use of the GR1 genomic clone N28 (section 4.2.6) to identify mis-incorporated nucleotides in the cDNA clone pGR165.

This analysis was simplified by the absence of any large intervening sequences in the region of the GR1 locus represented by the cDNA clone pGR165 (section 4.2.8). The 2kb BamHI-SalI restriction fragment of N28 (fig. 4.10) was subcloned into pBSM13^{+}, linearised with BamHI and SalI (section 2.16), to give the plasmid pBS28-A (fig. 6.3). Various restriction fragments of this clone were then subcloned (section 2.16) into M13 vectors for DNA sequence determination using Sequenase (section 2.22.4).

The strategy adopted for this sequencing exercise (fig. 6.3) was designed to confirm the DNA sequence of the region of pGR165 containing the three large open reading frames (fig. 6.2). This strategy also allowed determination of the DNA sequence between the BamHI restriction site in N28 and the 5′ end of the region represented by pGR165.

Comparison of the DNA sequences determined for pBS28-A with that of the cDNA clone pGR165 revealed only a single base substitution. A T residue at position 230 in pGR165 (fig. 6.2) was replaced with a G residue in pBS28-A, resulting in a methionine residue replacing an isoleucine residue in the predicted amino acid sequence (fig. 6.4). Given the known

Figure 6.4. Partial nucleotide sequence of GR1 transcription unit and putative GR1 polypeptide.

The consensus nucleotide sequence of the sense strand of the region of the GR1 transcription unit represented by pBS28-A and pGR165 is shown, along with the amino acid sequence of the single open reading frame of this sense strand. Restriction enzyme recognition sequences for BamHI, KpnI, PstI, SacI, and StuI are shown. The broken line marks the 5′ terminus of the cDNA clone pGR165 (fig. 6.2). The single nucleotide difference between cDNA and genomic sequences at nucleotide 457 is boxed, as is the cytosine-rich element between nucleotides 644-721. Consensus splice acceptor sites (Breathnach and Chambon, 1981) at positions 1066, 1912, and 1926 and AT-rich elements (Cosman, 1987) at positions 2004, 2499, 2528 and 2597 are boxed.

```
  BamHI                  20                            40
G GAT CCA CAC CTG CAT CAG CAC CAG CAT CCG CAG CCG CAG CCG CAT CCA CAA CAG
  Asp Pro His Leu His Gln His Gln His Pro Gln Pro Gln Pro His Pro Gln Gln

    60                           80                           100
  CAT CCT AAC CAT CCT CAG CAG CCA CCA TTC TTC TAC CAC ATG GCT GGA TTC AGA
  His Pro Asn His Pro Gln Gln Pro Pro Phe Phe Tyr His MET Ala Gly Phe Arg

           120                          140                          160
  ATT TAC CAC CCT GTA AGG TAT TAC TAT ATT CAG AAT GTG TAT ACA CCT GTT GAT
  Ile Tyr His Pro Val Arg Tyr Tyr Tyr Ile Gln Asn Val Tyr Thr Pro Val Asp

                     180                          200
  GAG CAT GTC TAT CCG GGT CAC CGG GTG GTT GAC CCT AAC ATT GAG ATG ATT CCT
  Glu His Val Tyr Pro Gly His Arg Val Val Asp Pro Asn Ile Glu MET Ile Pro

  220                  |   240                          260
  GGA GCG CAC AGC CTG CCC AGT GGA CAT TTG TAC TCA ATG TCT GAG TCT GAA ATG
  Gly Ala His Ser Leu Pro Ser Gly His Leu Tyr Ser MET Ser Glu Ser Glu MET
                       |

        280                          300                          320
  AAT GCT CTG CGA AAT TTC GTG GAC AGG AAT GTT AAA GAT GGG CTC ATG ACT CCC
  Asn Ala Leu Arg Asn Phe Val Asp Arg Asn Val Lys Asp Gly Leu MET Thr Pro

                  340                          360                          380
  ACT GTG GCG CCC AAT GGA GCC CAA GTC CTG CAA GTG AAA AGA GGG TGG AAA CTC
  Thr Val Ala Pro Asn Gly Ala Gln Val Leu Gln Val Lys Arg Gly Trp Lys Leu

                         400 SacI                     420
  CAA GTC ACT TAC AAT TGC CGA GCT CCA CAG AGT GGC ACC ATC CAA AAT CAG TAC
  Gln Val Thr Tyr Asn Cys Arg Ala Pro Gln Ser Gly Thr Ile Gln Asn Gln Tyr

       440                       T 460                          480
  CTA CGC ATG TCT CTT CCA AAT ATG GGA GAC CCT GCA CAC CTG GCA AGC TAT GGT
  Leu Arg MET Ser Leu Pro Asn Met Gly Asp Pro Ala His Leu Ala Ser Tyr Gly

                500                          520                          540
  GAA TTT GTC CAA GTT CCT GGC TAC CCA TAT CCA GCC TAT GTT TAC TAT ACA AGC
  Glu Phe Val Gln Val Pro Gly Tyr Pro Tyr Pro Ala Tyr Val Tyr Tyr Thr Ser

                        560  KpnI                    580
  CCG CAT ATG ATG ACT GCG TGG TAC CCA GTA GGA CGA GAT GTA CAT GGA CGA ATA
  Pro His MET MET Thr Ala Trp Tyr Pro Val Gly Arg Asp Val His Gly Arg Ile

    600                          620                          640
  ATC GTT GTG CCT GTT GTA ATC ACC TGG TCT CAA AAT ACG AAC CGC CAG CCT CCG
  Ile Val Val Pro Val Val Ile Thr Trp Ser Gln Asn Thr Asn Arg Gln Pro Pro

             660                          680                          700
  GTG CCC CAG TAT CCT CCT CCG CAG CCA CCT CCA CCA CCA CCA CCA CCT CCA CCG
  Val Pro Gln Tyr Pro Pro Pro Gln Pro Pro Pro Pro Pro Pro Pro Pro Pro

                     720                PstI     740                  760
  CCA CCA CCA CCT CCA CCA GCA TCA TCC TGC AGT GCT GCG TAG AACCTGTCATGTCCT
  Pro Pro Pro Pro Pro Pro Ala Ser Ser Cys Ser Ala Ala
```

780 800 820
TTGTAGTCTCTGCCCTCAACTTGATCCTGTGCAGCTTCTCAATCTATGACTGTGTGGTACTGGACCTTCAG

840 SacI 860 880 StuI 900
AGGCGCACAGAGCTCAAGTCAGTTTTCGTCTTGACTGCCACTTTATAAGTTGACAGGCCTGGGTTTTACTT

920 940 960
GTTAAAACCTCTCACCATCTCAATCACAGGCTGCCAAGTGTCTTTACAAAGAAGCTGATACAAACACAGGC

980 1000 1020 1040
CATGCTGATTTCTTACAGAGGGAGAGAAGAGGAAGAGAAGAAGAAAGAGGAGGAAGAGGACATGACTTGCC

1060 1080 1100
CATATGCTGGGCACCTTATAAAGGAAGCCAGACTTTTCGGTGCAGTATGGAAAGGCTTCCGTGATTCTCTT

1120 1140 1160 1180
GCTGCACCCCACGAAACTTCACCACCTTCAAACTCCATTTTCACGGTTCCGTTAATTTTCAAGGAGCAGCA

1200 1220 1240
ACTCGACTGGTTCTCTGCTACATGAAACACCTCAGCTTGAAAAGGAAGTGCTCTCTCAGACTGACTTGTGA

1260 1280 1300 1320
GTGTGCCTTCACATTCTGGTGCAAATCATGTGTACCCAAGAACTCTGACATAGCATCTTACCATCATCATG

1340 1360 1380 1400
CCAGGAATTGGTTTCCTTAAAGTGTGACACTTGGCAACCAGAAGCGTCTGGTGGCTTCAGGTGTGTTAGGG

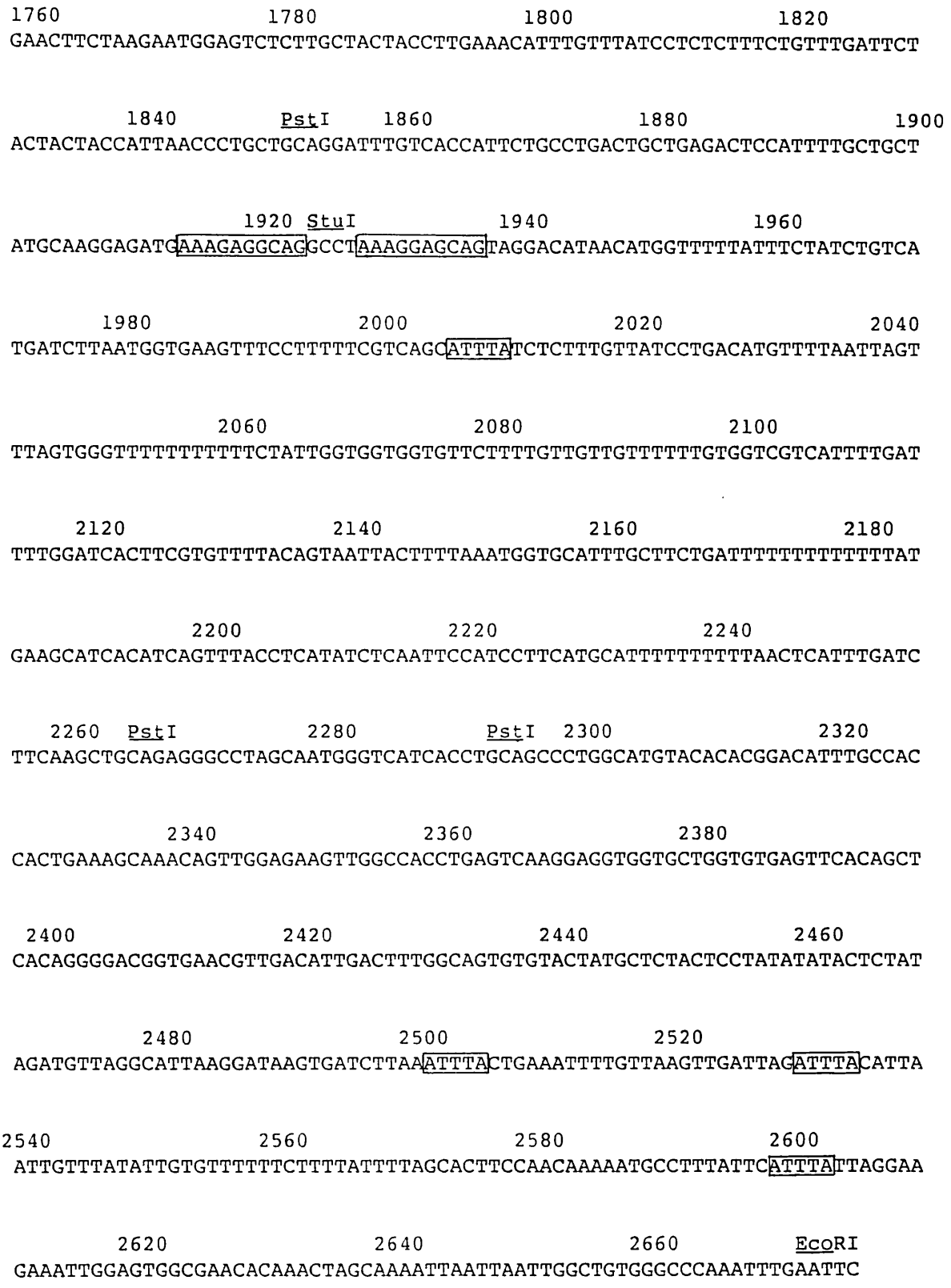
1420 1440 1460 KpnI
TTTTGTAGAAAGTTGTAACCACCTCTCAGTCATGCAAGAGGGTACATCAAAAAGACTTCTAAATTTGGTAC

1480 1500 1520 1540
CCAGACTTTGACAAGAGCCTGTGAGTTAGTTCATTGTAACCAACTTGTCAGGAACTTCAGCAAGAATTTTCAT

1560 1580 1600
GAATGTTGTTAATTGGCTGGAGAAATACCAGCAAAGTTTTGCAATCATCCACCCATTTGTCTGCCATTGTG

1620 1640 1660 1680
CATCTGGACATCCATCAACCGCCCACTCAGTGGCATCTACCATGAAGATGTTTAGAGACGGGAAATGATTT

1700 1720 1740
TGCACCCAGATCCACACTCCTCTCATGTCAGATATTTGGAATTAGAGCACCGGGAATCGCCTTGAAGATAT


1760
1780
1800
1820
GAACTTCTAAGAATGGAGTCTCTTGCTACTACCTTGAAACATTTGTTTATCCTCTCTTTCTGTTTGATTCT
1840
PstI
1860
1880
1900
ACTACTACCATTAACCCTGCTGCAGGATTTGTCACCATTCTGCCTGACTGCTGAGACTCCATTTTGCTGCT
1920
StuI
1940
1960
ATGCAAGGAGATGAAAGAGGCAGGCCTAAAGGAGCAGTAGGACATAACATGGTTTTTATTTCTATCTGTCA
1980
2000
2020
2040
TGATCTTAATGGTGAAGTTTCCTTTTTCGTCAGCATTTATCTCTTTGTTATCCTGACATGTTTTAATTAGT
2060
2080
2100
TTAGTGGGTTTTTTTTTTTCTATTGGTGGTGGTGTTCTTTTGTTGTTGTTTTTTGTGGTCGTCATTTTGAT
2120
2140
2160
2180
TTTGGATCACTTCGTGTTTTACAGTAATTACTTTTAAATGGTGCATTTGCTTCTGATTTTTTTTTTTTTAT
2200
2220
2240
GAAGCATCACATCAGTTTACCTCATATCTCAATTCCATCCTTCATGCATTTTTTTTTTAACTCATTTGATC
2260
PstI
2280
PstI
2300
2320
TTCAAGCTGCAGAGGGCCTAGCAATGGGTCATCACCTGCAGCCCTGGCATGTACACACGGACATTTGCCAC
2340
2360
2380
CACTGAAAGCAAACAGTTGGAGAAGTTGGCCACCTGAGTCAAGGAGGTGGTGCTGGTGTGAGTTCACAGCT
2400
2420
2440
2460
CACAGGGGACGGTGAACGTTGACATTGACTTTGGCAGTGTGTACTATGCTCTACTCCTATATATACTCTAT
2480
2500
2520
AGATGTTAGGCATTAAGGATAAGTGATCTTAAATTTACTGAAATTTTGTTAAGTTGATTAGATTTACATTA
2540
2560
2580
2600
ATTGTTTATATTGTGTTTTTTCTTTTATTTTAGCACTTCCAACAAAAATGCCTTTATTCATTTATTAGGAA
2620
2640
2660
EcoRI
GAAATTGGAGTGGCGAACACAAACTAGCAAAATTAATTAATTGGCTGTGGGCCCAAATTTGAATTC

error rate of AMV reverse transcriptase during cDNA synthesis (section 6.1.1), the genomic DNA sequence was used in subsequent analyses. No other nucleotide differences were found between pGR165 and the regions of the GR1 genomic clone N28 analysed (data not shown). The additional 237 nucleotides of pBS28-A not present in pGR165 were added to the consensus sequence of pGR165 to give the partial DNA sequence of the GR1 transcription unit shown in figure 6.4. Analysis of this sequence extended the first open reading frame of pGR165, between 6-515 (fig. 6.2), a further 77 amino acids towards the presumed amino terminus of the putative GR1 polypeptide. No other large open reading frame was found in the sequence unique to pBS28-A (data not shown).

Thus, an open reading frame, with the potential to encode 247 amino acids, was found to extend from the BamHI restriction site of pBS28-A to just beyond the PstI restriction site at position 734 (fig. 6.4). The termination codon of the putative GR1 polypeptide is provided by a TAG codon at position 743 (fig. 6.4), there being 1.93kb of potential 3′ untranslated sequences in the 2674 nucleotides of the GR1 locus determined (fig. 6.4). In the absence of any 5′ untranslated DNA sequences, the predicted amino acid sequence, of molecular weight 27.9 kilodaltons, represents only the carboxy terminus of the putative GR1 polypeptide.

Section 6.2.4. Analysis of the partial DNA sequence of the GR1 transcription unit.

The nucleotide sequence of the GR1 locus determined in section 6.2.4 was subjected to a number of computer-assisted analyses. Comparison of this DNA sequence with all those in the October 1987 release of the EMBL data base revealed no significant homologies. Using the Wilbur and Lipman algorithm (Wilbur and Lipman, 1983), with k-tuple =4, the strongest homology detected was 29.2% with Epstein Barr virus genome sequences 48047-50717 (data not shown). Such homology is no greater than that expected by chance.

Recognition sequences for all known restriction enzymes sites in pGR165 (fig. 6.1) and pBS28-A (fig. 6.3) were present in the determined sequence. An additional StuI recognition sequence was detected at position 886 which was not present in pGR165. However, the three G residues following this recognition sequence would make it susceptible to dcm methylation (section 4.3.1). Thus, the sequence data provide an explanation for the apparent anomaly between restriction maps of the GR1 locus and the cDNA clone pGR165 (section 4.2.1). Three Sau3a recognition sequences were located between nucleotides 1971-2250 in the GR1 nucleotide sequence. One of these is likely to represent the cloning site of the genomic clone N28 (section 4.2.6).

Analysis of the putative 3′ untranslated region of the determined sequence revealed the presence of 4 copies of the

consensus sequence ATTTA (fig. 6.4) thought to confer instability on many lymphokine, cytokine and proto-oncogene mRNA species (Cosman, 1987). These elements were found within AT rich tracts of the putative 3′ untranslated region of the GR1 mRNA (fig. 6.4). No polyadenylation signals (Breathnach and Chambon, 1981) or poly A tail were found in the putative 3′ untranslated region.

There was no evidence that the sequence determined contained any intervening sequences. No consensus splice donor sites (Breathnach and Chambon, 1981) were found, and a single open reading frame extended from the 5′ end of the determined sequence to position 743 in the nucleotide sequence (fig. 6.4). However, the presence of three consensus splice acceptor sites (Breathnach and Chambon, 1981) in the 3′ untranslated region of the determined sequence (fig. 6.4) leaves open the possibility that differential splicing of the primary GR1 transcript may occur.

Section 6.2.5. Analyses of putative GR1-encoded polypeptide.

All three of the large open reading frames in the determined sequence of the sense strand of the GR1 locus (fig. 6.2; fig. 6.4) were compared with the 5205 amino acid sequences in the July 1987 release of the EMBL data base. Comparisons were carried out using the Wilbur and Lipman algorithm with a k-tuple value of 2.

Figure 6.5. Amino acid composition of putative GR1 polypeptide.

The amino acid composition of the region of the putative GR1 polypeptide determined by nucleotide sequence analysis (fig. 6.4) is compared with the average amino acid composition of all known eukaryotic proteins (Doolittle, 1986).

Figure 6.5. Amino acid composition of putative GR1 polypeptide.

Amino Acid	Percentage GR1	Percentage All Eukaryotes	Amino Acid	Percentage GR1	Percentage All Eukaryotes
Asp	2.82	5.4	Met	3.63	2.0
Asn	4.84	5.4	Val	8.87	6.1
Thr	3.63	5.8	Ile	3.23	6.1
Ser	5.24	7.1	Leu	4.03	8.9
Glu	2.02	5.8	Tyr	7.26	3.8
Gln	8.06	3.9	Phe	2.02	4.3
Pro	19.35	4.6	Trp	1.21	1.4
Gly	5.24	6.9	His	6.45	2.4
Ala	5.24	7.5	Lys	1.21	6.5
Cys	0.81	1.7	Arg	4.44	4.3

The best homology of the putative GR1-encoded polypeptide sequence (fig. 6.4) was found to be 22.2% with that of the FGR tyrosine kinase transforming protein (data not shown). This homology was evenly distributed. The two additional open reading frames, from positions 263-649 and 658-951 of the sequence of pGR165 (fig.6.2), had best homologies with a merozoite surface antigen precursor (28.8%) and human prolactin precursor sequences, respectively (data not shown). From such low homologies, it was concluded that GR1 has the potential to encode a novel polypeptide.

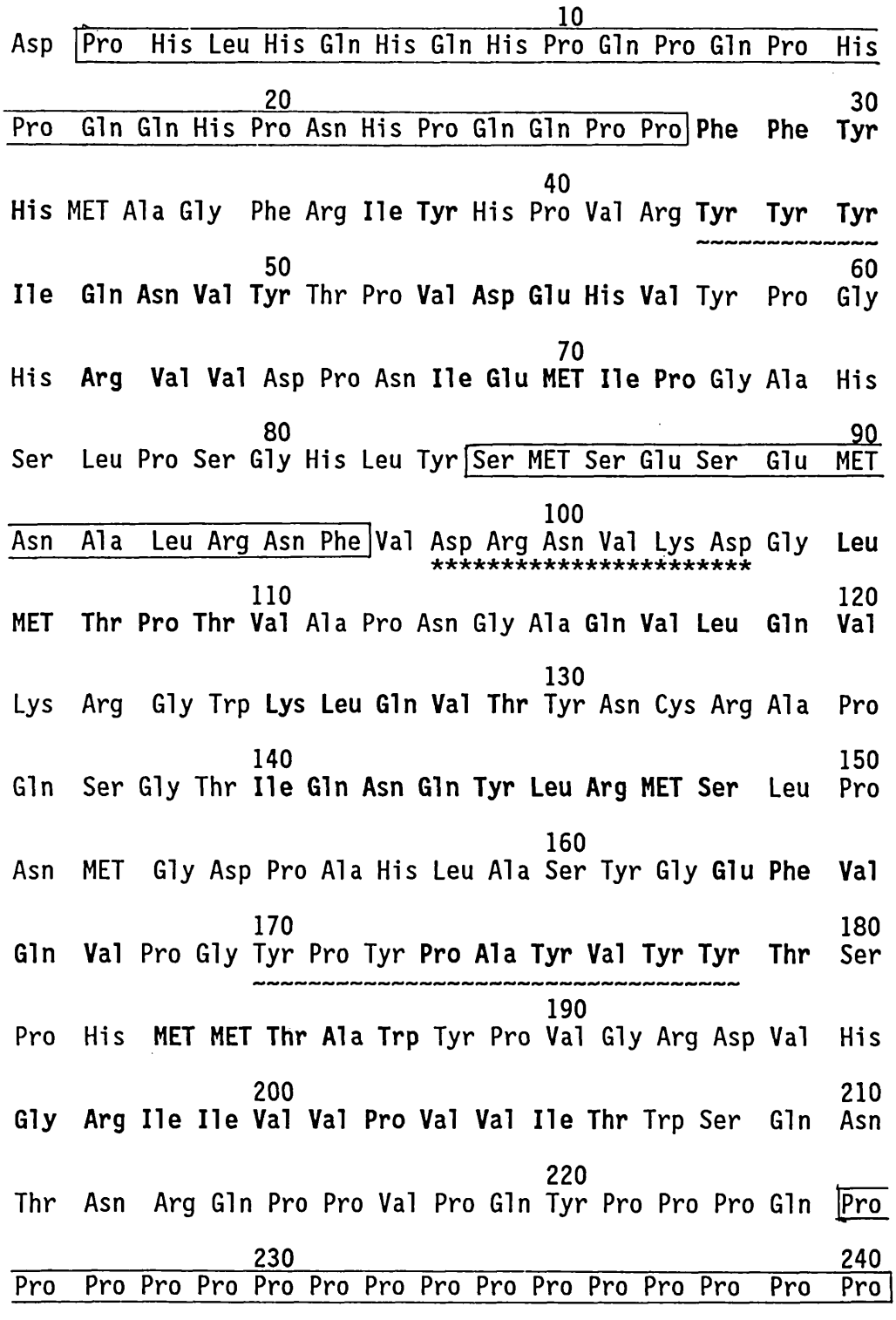
Analysis of the amino acid composition of the putative GR1-encoded polypeptide (fig. 6.4) revealed this sequence to be rich in proline, glutamine, tyrosine, and histidine, relative to the average amino acid composition of known eukaryotic proteins (Doolittle, 1986; fig. 6.5). Moreover, these residues were found to be clustered in the predicted amino acid sequence, most notably the proline-rich region towards the carboxy terminus of the putative polypeptide (fig. 6.6). Additionally, a Gln, His, Pro rich region was located between residues 2-27 (fig. 6.6). Tyrosine residues were also seen to be clustered, for example at residues 43-45 and 170-178 (fig. 6.6).

Analysis of the putative GR1 amino acid sequence by use of hydropathy plots (Kyte and Doolittle, 1982), and the Robson secondary structure prediction programme (Garnier et al., 1978), enabled predictions to be made concerning the structure of this novel polypeptide (fig. 6.6).

Figure 6.6. Secondary structure predictions of putative GR1 polypeptide.

Secondary structures of the partial amino acid sequence of the putative GR1 polypeptide, predicted by the Robson method (Garnier et al., 1978) are shown. **Boldtype** represents β-sheet structures. Boxed regions represent the His, Gln, Pro rich (residues 2-27) and poly-proline (residues 225-240) regions, both of which are predicted to assume unique coiled conformations (section 6.2.5), and the single α-helix between residues 84-96. ~~~~= Tyrosine-rich regions. **** = Most likely antigenic site (Hopp and Woods, 1981).

Figure 6.6. Secondary structure predictions of putative GR1 polypeptide.



From such analyses, it was found that this region of the putative GR1 polypeptide is rich in β-sheet structure, the length of which is approximately that expected of a typical protein domain (W. Taylor, pers. comm.). The β-rich domain is divided in two by a single prominent α-helix, between residues 84-96 (fig. 6.6). Within the β-sheets, proline residues are spaced in a periodic manner and may serve to introduce turns into the β-sheets.

This β-rich domain is flanked by coiled structures. On the amino terminal side is the Gln, His, Pro rich region which, from the properties of such residues, is predicted to assume a rigid coiled conformation (W. Taylor, pers. comm.). Similarly, the poly-proline region is predicted to form a coiled structure. From X-ray diffraction analysis of synthetic poly-L-proline (Cowan and McGavin, 1955), it is known that this poly amino acid can assume a unique helical conformation, stabilised by steric repulsion of the pyrollidone rings in the proline residues. The poly-proline region in the putative GR1 polypeptide could represent 5 turns of such a helix.

Beyond the poly-proline region lies the final 7 residues, containing serine and alanine residues and one of only two cysteine residues in the 247 amino acid sequence (fig. 6.6). Analysis of the putative GR1 amino acid sequence failed to reveal any sequences with homology to the active sites of serine proteases, acid proteases, hydroxy-kinases, and a number of other enzymes (Doolittle, 1986). In addition, no sequences homologous to Asn-linked carbohydrate attachment

sites, nucleotide binding domains, metal binding domains, or cell adhesion sites (Doolittle, 1986) were found (data not shown).

Whilst a strongly hydrophobic domain was found between residues 198-207 (fig. 6.6), this was not large enough to represent a hydrophobic transmembrane domain, characteristically 20-30 residues in length (Doolittle, 1986). From analysis of the hydrophilic character of the putative GR1 amino acid sequence (Hopp and Woods, 1981), it was predicted that the peptide Asp-Arg-Asn-Val-Asp-Lys would represent the most likely antigenic site. This sequence forms part of the predicted turn on the carboxy terminal side of the prominent α-helical region in the predicted secondary structure of the putative GR1-encoded sequence (fig. 6.6).

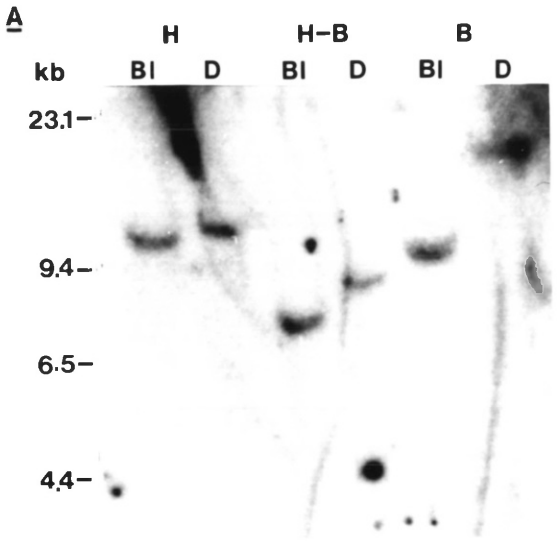
Section 6.2.6. Mapping the GR1 locus in BxD recombinant inbred mice.

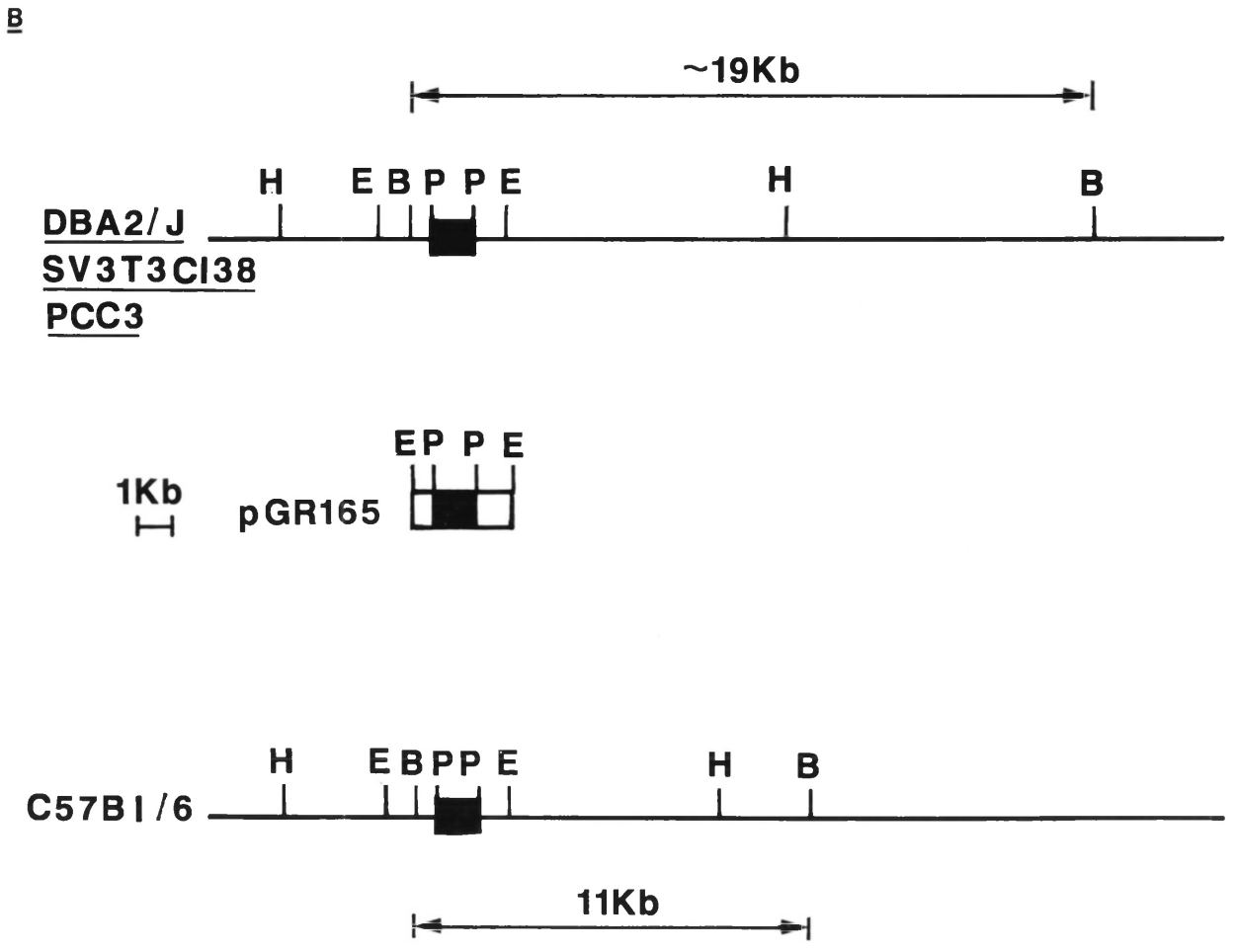
As discussed earlier (section 6.1.3), in order to utilise recombinant inbred strains for gene mapping studies it was necessary that the alleles of the locus of interest in the two progenitor strains could be distinguished. For a novel gene, such as GR1, this was most easily achieved by the use of a DNA restriction fragment length polymorphism (RFLP) at the locus under investigation. Thus, in order to establish the strain distribution pattern of the GR1 locus in the BxD

Figure 6.7. Restriction fragment length polymorphisms at GR1 locus.

A. Genomic DNA prepared from livers of DBA2/J (D) and C57bl/6 (Bl) mice was digested with HindIII (H), BamHI (B) or HindIII and BamHI (H-B), electrophoresed through a 0.4% (w/v) agarose gel, the fragments transferred to a nitrocellulose filter (section 2.18) and probed with radiolabelled RNA synthesised *in vitro* using T3 polymerase on a template of pBS1200A linearised with SacI (specific activity $7.6x10^8$cpm/ug, 10^6cpm/ml used). The filter was washed to 0.1xSSC, 0.1% (w/v) SDS at 65^oC and exposed for 2 days. The size of hybridising restriction fragments was determined by comparison with the rate of migration of restriction fragments of bacteriophage lambda DNA digested with HindIII. kb= kilobase.

B. Restriction map showing HindIII and BamHI polymorphisms at GR1 locus between DBA2/J and C57bl/6 mouse strains based on the data in A. The black box designates the region of pGR165 used as probe and the region of the GR1 locus homologous to this probe. The restriction map of the GR1 locus in DBA2/J mice is the same as that in the murine cell lines PCC3 and SV3T3 C138 (fig. 4.3), cell lines derived from 129/SV and Balb/c mouse strains, respectively. B= BamHI, E= EcoRI, H= HindIII, P= PstI.


A
H
H-B
B
kb
Bl
D
Bl
D
Bl
D
23.1
9.4
6.5
4.4


B
~19Kb
H
E
B
P
P
E
H
B
DBA2/J
SV3T3Cl38
PCC3
1Kb
pGR165
E
P
P
E
H
E
B
P
P
E
H
B
C57Bl/6
11Kb

recombinant inbred strains, it was necessary to detect an RFLP at this locus between DBA2/J and C57bl/6 strains of mice.

On the assumption that GR1 has an essential role in the mouse life cycle, no RFLPs were expected for restriction sites within known coding sequences of this locus. Hence, attempts to detect polymorphisms at the GR1 locus were limited to analysis of restriction enzyme sites lying outside the known transcribed region of GR1 represented by the cDNA clone pGR165 (fig. 4.3). The 1.1kb PstI-PstI restriction fragment of pGR165 was used to generate radiolabelled probes specific for the GR1 locus (fig. 4.9). Initial Southern blotting experiments using restriction fragments of genomic DNA prepared from either DBA2/J or C57bl/6 mice failed to detect any polymorphisms, at the GR1 locus, for EcoRI, KpnI or SacI (data not shown). However, RFLPs at the GR1 locus between DBA2/J and C57bl/6 were detected for restriction enzymes HindIII and BamHI (fig. 6.7a). Additional Southern blot analyses located both polymorphic restriction enzyme sites to the 3′ end of the GR1 locus, with regard to the transcriptional orientation of this gene (fig. 6.7b).

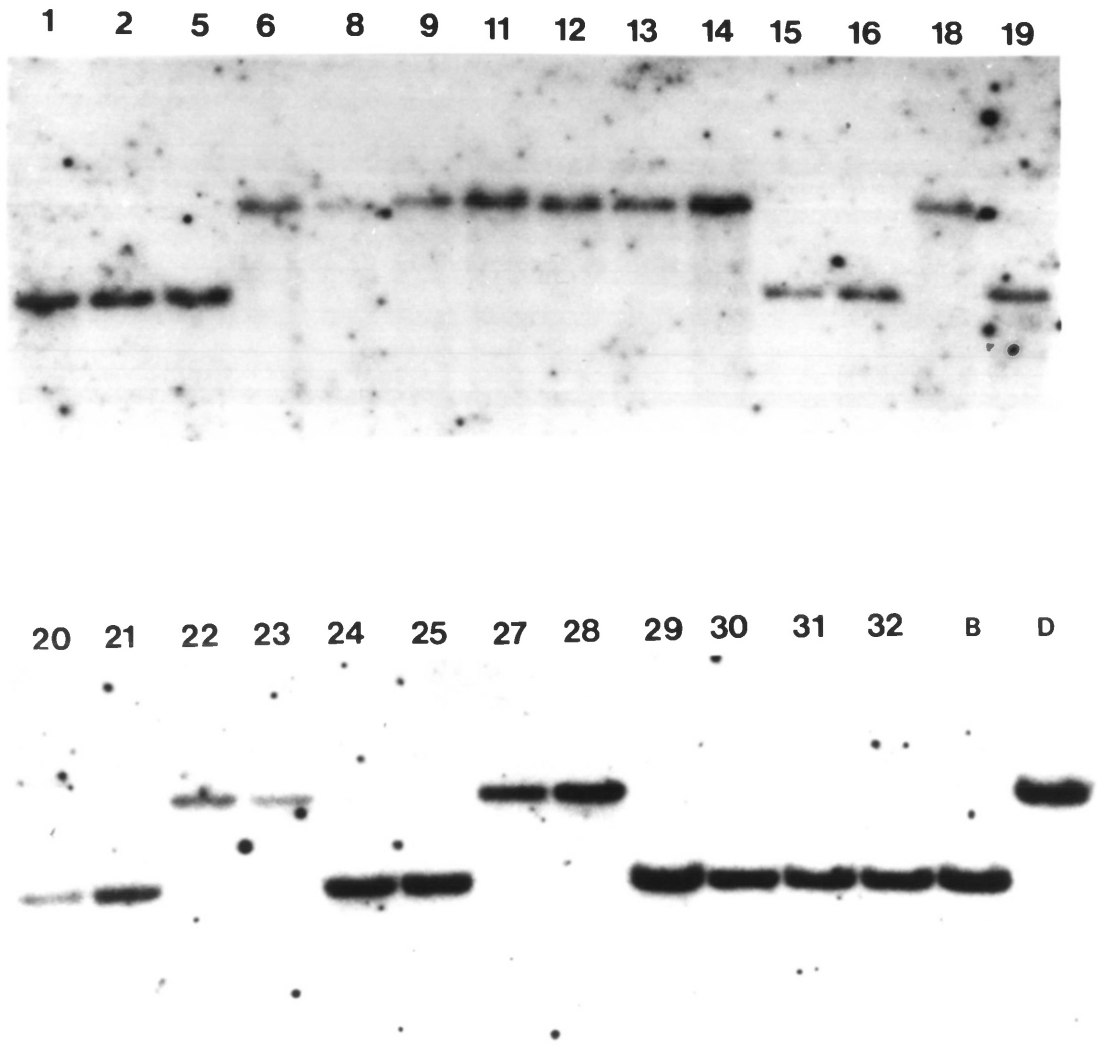
The BamHI polymorphic site was then used to determine the strain distribution pattern (SDP) of the GR1 locus in Southern blotting experiments of BamHI restriction fragments of genomic DNA isolated from each of 26 recombinant inbred strains in the BxD series. Clear segregation of this polymorphism was seen in each of the strains analysed (fig. 6.8a). Computer assisted comparisons of the SDP of the GR1 locus with 186 other mouse

Figure 6.8. Strain distribution pattern of the GR1 locus in BxD recombinant inbred mice.

A. A nitrocellulose filter containing electrophoresed BamHI restriction fragments of genomic DNA from a number of the BxD recombinant inbred mouse strains (1-32), C57bl/6 (B) and DBA2/J (D) mice (the kind gift of Dr Bernhard Herrmann) was probed with the 1000bp SacI-PstI fragment of pBS1200A (fig. 3.5) radiolabelled by the method of random priming (section 2.17) (specific activity $2.4x10^9$cpm/ug, $1.6x10^6$cpm/ml used). Both filters were washed to 0.1xSSC 0.1%(w/v) SDS at 65^oC and exposed for 2 days.

B. The SDP of the proto-oncogene met locus (Dean et al., 1987) in BxD recombinant inbred strains of mice is compared to that of the GR1 locus. *= recombinants.

A



B

Loci	BxD Strain																									
	01	02	05	06	08	09	11	12	13	14	15	16	18	19	20	21	22	23	24	25	27	28	29	30	31	32
GR1	B	B	B	D	D	D	D	D	D	D	B	B*	D	B	B	B	D	D	B*	D	B	B	B	B	B	B
met	B	B	B	D	D	D	D	D	D	D	B	D	D	B	B	B	D	D	D	B	D	B	D	B	D	D

loci mapped in the BxD series, showed GR1 to be tightly linked with the locus of the proto-oncogene *met* (fig. 6.8b; B. Taylor, pers. comm.). No tight linkage was detected for any other locus mapped in this series of recombinant inbred strains (B. Taylor, pers. comm.). Using the calculation of Taylor *et al.* (1975), the recombination frequency between GR1 and *met* was calculated as 0.05. From these data, it was estimated that GR1 lies between 1.17cM and 18.28cM from the *met* locus at the 95% confidence interval (Silver, 1985). Given the limited data available, it was not possible to determine whether GR1 is proximal or distal to *met*.

Section 6.3. Discussion.

Section 6.3.1. Nucleotide sequence analyses of GR1 transcription unit.

Using both cDNA and genomic clones of the GR1 locus, the sequence of 2674 nucleotides of the GR1 transcription unit was determined. No intervening sequences were found in this region. This, along with the finding that the StuI restriction enzyme recognition sequence at nucleotide position 886 (fig. 6.4) is a potential substrate for dcm methylation (section 4.3.1), confirms the conclusion from restriction mapping data (section 4.3.1) that the region characterised represents a single exon of the GR1 locus. Neither 5′ nor 3′ limits of this exon were determined.

The inability to isolate any cDNA clones containing sequences transcriptionally upstream of those represented by pGR165 may reflect some stable secondary structure of the GR1 mRNA which makes this region unsuitable as a template for AMV reverse transcriptase. Indeed, a putative stem-loop configuration involving 3 stable G-C base pairs can be formed at the site of termination of the cDNA clone pGR165 (fig. 6.9a). None of the cDNA clones analysed were found to contain rearranged sequences, suggesting that instability of GR1 cDNA clones propagated in E.coli was not a problem.

A single open reading frame in the sense-strand of pGR165 was found to terminate at nucleotide 515 (fig. 6.2). However, the

Figure 6.9.

A. Potential stem-loop structure at termination site of pGR165.

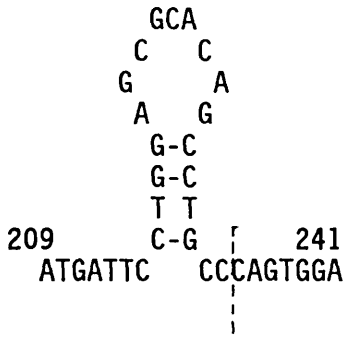
The potential secondary structure of nucleotides 209-241 of the GR1 DNA sequence (fig. 6.6) is shown. This structure is stabilised by three G-C base pairs. The broken line designates the termination site of pGR165.

B. Open reading frames of pUR-12 fusion proteins.

The predicted amino acid sequences across the PstI restriction enzyme site used to construct the pUR-12 series of β-galactosidase-GR1 fusion proteins (section 5.2.2) are shown. The 3 carboxy terminal amino acids of the putative GR1 polypeptide (fig. 6.6) are found in 290-12. Fusion proteins 291-12 and 292-12 are predicted to contain 47 and 23 amino acids, respectively, of each of the other two potential reading frames of pGR165. The nucleotide sequence of pUR vectors is given in Ruther and Muller-Hill (1983). The nucleotide sequence of pGR165 (boldface) was derived from fig. 6.4.

Figure 6.9.

A. Potential stem-loop structure at termination site of pGR165.



B. Open reading frames of pUR-12 fusion proteins.

```
                                    PstI
290-12  ....GGG GAT CCG TCG ACC TGC AGT GCT GCG TAG
            Gly Asp Pro Ser Thr Cys Ser Ala Ala  :

                                    PstI
291-12  ...GG GGA TCC GTC GAC CTG CAG TGC TGC GTA GAA...47AA.
              Gly Ser Val Asp Leu Gln Cys Cys Val Glu.....

                                    PstI
292-12  ...G GGG ATC CGT CGA CCT GCA GTG CTG CGT AGA..23AA.
            Gly Ile Arg Arg Pro Ala Val Leu Arg Arg.....
```

presence of two additional large open reading frames in this sense strand of pGR165 (fig. 6.2) raised the possibility that small deletions, insertions or substitutions during reverse transcription had altered the nucleotide sequence of the GR1 transcription unit.

Determination of the nucleotide sequence of equivalent regions of the GR1 locus, represented by the genomic subclone pBS28-A (fig. 6.3), resulted in only a single nucleotide difference being found between cDNA and genomic sequences. The substitution of T in the cDNA clone pGR165, in place of G at nucleotide 457 in the genomic DNA sequence, altered the predicted amino acid residue at this position from Ile to Met (fig. 6.4). No other differences were found between cDNA and genomic nucleotide sequences. Whilst there were 4 short regions, within the putative open reading frames, for which genomic sequence was determined for only one strand (fig. 6.3), these sequences were clearly legible and, when applicable, identical with that derived previously for the cDNA clone pGR165.

Further confirmation that the first open reading frame in pGR165 was correct came from the finding that the single open reading frame of the additional 237 nucleotides of the GR1 sense strand, unique to the genomic subclone pBS28-A, was in the same frame as that of the cDNA clone (fig. 6.4).

Establishment of the open reading frame of the cDNA clone pGR165 provides an explanation for the negative results of experiments designed to generate anti-GR1 antisera from the

fusion protein 291-12 (section 5.2.3). From the sequence data presented here, it is seen that the 291-12 fusion protein contains 47 amino acids of the second open reading frame of pGR165 (fig. 6.2; fig. 6.9b). The estimated size of the 291-12 fusion protein (section 5.2.2) is close to that predicted from sequence data.

The predicted GR1 polypeptide-encoding frame is represented by fusion protein 290-12 (fig. 6.9b). However, this latter fusion protein would contain only 3 amino acid residues additional to native β-galactosidase and so, given the criteria used in the choice of fusion protein (section 5.1.2), was not used for immunisations. The third reading frame of this region of pGR165 is represented by fusion protein 292-12, and is predicted to contain an additional 23 amino acids fused with β-galactosidase. It is unclear why the rate of migration of this fusion protein was the same as native β-galactosidase on SDS-PAGE (section 5.2.2). The fusion protein 292-12 may have been unstable and degraded to a stable β-galactosidase sized molecule.

Within the putative 3′ untranslated region of the GR1 locus, four copies of the sequence ATTTA were found within A-T rich regions (fig. 6.4). Such an organisation is similar to the 3′ untranslated regions of many cytokine, lymphokine and proto-oncogene mRNA species (Cosman, 1987). The common feature of these genes is that, under certain circumstances, the corresponding mRNA species have short half lives (Cosman, 1987).

A synthetic nucleotide sequence, corresponding to the AT rich region of human GM-CSF, was found to confer instability on β-globin mRNA when inserted into the 3′ untranslated region of this gene (Shaw and Kamen, 1986). Deletion of sequences, including AT-rich consensus sequences, in the 3′ untranslated region of c-fos, correlates with the acquisition of transforming properties of this oncogene (Meijlink et al., 1985). Deletion of the same sequences has been shown to stabilise c-fos mRNA in both normal and mitogen-stimulated fibroblasts (Treisman, 1985). More recently, it has been suggested that it may be the secondary structure of such AT-rich regions which is responsible for mRNA instability (Jones and Cole, 1987).

It may be that, under certain circumstances, the AT-rich elements in the 3′ untranslated region of GR1 are involved in the regulation of this gene at the level of RNA stability. Indeed, the observed decrease in steady-state levels of GR1 mRNA in differentiating PCC3 cells (fig. 3.11) is greater than that seen in transcriptional run-on analyses, suggesting that there may be a post-transcriptional component to the regulation of GR1 expression (Reith and Cooper, unpublished observations).

Any role for the AT-rich elements in the regulation of GR1 expression could be established by analysis of expression of plasmid constructs, containing the GR1 gene under the regulation of its own promoter elements, but deleted in the AT rich domains. The termination of the GR1 genomic clone, N28,

within this region (section 6.2.4) may provide a suitable deletion construct for preliminary analysis, once the transcriptional regulatory elements of this gene have been characterised.

No polyadenylation signal nor poly A tail was detected in the putative 3′ untranslated regions of the GR1 transcription unit characterised (fig. 6.4). This strengthens the tentative conclusion from restriction mapping of the GR1 locus (section 4.3.1) that the EcoRI restriction site at the 3′ end of pGR165 is part of the GR1 transcription unit and that the sequences characterised may represent regions of the GR1 transcription unit some way upstream of the 3′end of this gene. The presence of 3 consensus splice acceptor sites in the 3′ untranslated region of GR1 suggests that differential splicing of a primary GR1 transcript may occur. Whilst the 5′ limit of the GR1 exon characterised in the present work is not known, the splice acceptor sites detected could be used to splice with exons upstream of that represented by pGR165. Such splicing could result in the generation of GR1 mRNA species with different polypeptide-encoding capacities but sharing common 3′untranslated sequences. Evidence for such differential splicing could come from DNA sequence analysis of cDNA clones with homology to regions of pGR165 downstream from the putative splice acceptor sites.

The most striking feature of the determined nucleotide sequence is the cytosine-rich region located between nucleotides 644-721 (fig. 6.4). Whilst non-specific

terminations prevented determination of the sense strand of this sequence using Klenow fragment of E.coli DNA polymerase I (section 2.22.3), that of the lower strand was clearly resolved. Both strands of this region were readily determined using the modified T7 DNA polymerase, Sequenase (section 2.22.4). This cytosine-rich element encodes a proline-rich domain located towards the carboxy terminus of the putative GR1 polypeptide (fig. 6.6).

This cytosine-rich region lies within the KpnI-PstI restriction enzyme fragment previously characterised as containing a moderately repetitive element in the mouse genome (section 4.2.2). Given that the cytosine-rich element is homologous to a number of nucleotide sequences on the EMBL data base, including mouse salivary proline rich protein (PRP) genes (Ann and Carlsson, 1985) and fibrous collagens, it is highly likely that it is this sequence which represents the repetitive element seen on Southern blot analysis of mouse genomic DNA. The G-rich complementary strand of this region is also homologous to several genes on the data base including keratin genes.

It would be of interest to determine whether the cosmid clones CAD18, CAD20 and CAD22 (section 4.2.4) correspond to any of these genes, or represent novel CG-rich genomic sequences. This could be achieved simply by Southern blot analyses, if appropriate probes were available. Should any of these cosmid clones represent novel genes, the amino acid encoding capacity of such CG rich clones could be investigated, since no protein

with such an extensive poly-proline region as that in GR1 is included in the current EMBL amino acid sequence data base. The relatively large number of CpG dinucleotides in the cytosine-rich element makes it likely that this region of the GR1 locus contains a high proportion of methylcytosine residues. As such, this element would be a particularly susceptible substrate for E.coli methylcytosine restriction systems (section 4.1.3), and provides a possible explanation for the isolation of GR1 genomic clones only when recombinant libraries were propagated on methylcytosine restriction deficient host E.coli strains (section 4.2.6).

Section 6.3.2. Analyses of putative GR1 polypeptide.

As discussed earlier (section 6.1.2), in the absence of general homology between the putative amino acid sequence of GR1 and other amino acid sequences, it is not possible to predict the tertiary structure of the GR1 polypeptide. The abundance of β-sheets in the predicted secondary structure of this region of the putative GR1 polypeptide (fig. 6.6) suggests that GR1 may encode an extracellular non-enzymatic protein (Nishikawa et al., 1983; section 6.1.2). Clearly further analysis is required to confirm this prediction. The derived amino acid sequence can also be used in the design of subsequent experiments to question the function of GR1.
In this respect, the peptide Asp-Arg-Asn-Val-Asp-Lys was

predicted as representing the most likely antigenic determinant in the the GR1 sequence obtained (fig. 6.6). For reasons discussed earlier (section 6.1.2), the internal location of this peptide may make it unsuitable for production of anti-GR1 antisera. Peptides of the Pro, His, Gln rich domain, at the amino terminus of the derived sequence, may prove to be more useful. The rigid coiled secondary structure predicted for this region (section 6.2.6) is likely to be maintained in synthetic peptides. The poor hydrophilic properties predicted for of the carboxy-terminal peptide Ala-Ser-Ser-Cys-Ala-Ala may make this peptide unsuitable as an antigen.

The nucleotide sequence could also be used as the basis for the construction of further GR1-fusion proteins in attempts to raise GR1-specific antisera. Such reagents, in addition to being vital in subsequent efforts to determine potential functions for this novel gene, would go a long way to confirming the amino acid sequence predicted here.

Antisera raised against the poly-proline region of the putative GR1 polypeptide would be of greater use than nucleic acid probes in the isolation of additional genes containing such proline rich regions. However, care would need to be exercised to accommodate potential cross-reactivity between anti-poly-proline antisera and other previously characterised proline-rich proteins.

The predicted conformation of the poly-proline helix is such as to facilitate hydrogen bonding, perhaps via hydroxyproline

residues, between different GR1 polypeptide chains or other proteins. However, investigation of any such interactions awaits purification of the GR1-encoded protein. This could be achieved given the availability of GR1-specific antisera.
The amino acid sequence presented here represents only the carboxy-terminal domain of the putative GR1 polypeptide. Clearly, further DNA sequence analysis is required in order to determine the complete amino acid sequence of polypeptide(s) encoded by this novel gene. In conjunction with transcriptional mapping analyses, the GR1 genomic clone, N28 (section 4.2.6), could be used for this purpose.

Section 6.3.3. Chromosome mapping of GR1 locus.

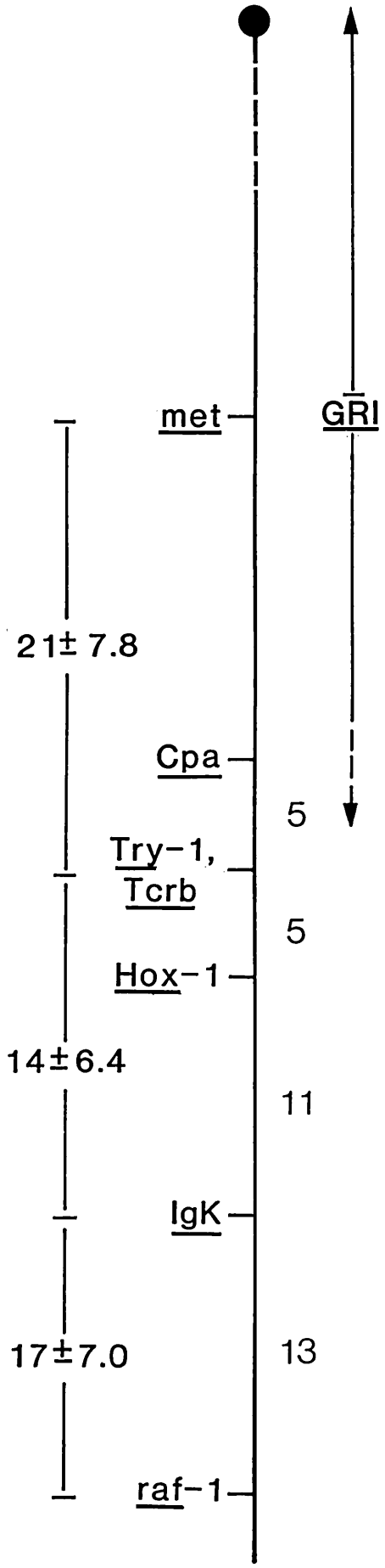
A BamHI restriction fragment length polymorphism (RFLP) at the 3′ end of the GR1 locus was used to determine the strain distribution pattern (SDP) of the GR1 locus in the BxD series of recombinant inbred mouse strains. Comparison of this distribution pattern with that of 186 other loci mapped in the mouse genome in this way, showed the GR1 locus to be tightly linked to, but not identical with, the proto-oncogene *met* (Dean *et al.*, 1987). The observed recombination frequency of 0.05 between GR1 and *met* corresponds to an estimated genetic distance of 1.17cM to 18.28cM, at a 95% confidence interval (Silver, 1985). It is not known if GR1 is proximal or distal to *met*.

On the basis of analysis of somatic cell hybrids and interspecies backcrosses, in addition to the SDP in the BxD and other recombinant inbred strains, the met locus has been determined to lie proximal to the T cell receptor β-chain (Tcrb) locus on mouse chromosome 6 (Dean et al., 1987; fig. 6.10).

Other than GR1, the locus closest to met for which the SDP in the BxD series is available is that of Hox-1.4 and Hox-1.7 (Bucan et al., 1986; Rubin et al., 1986; Martin et al., 1987; B.Taylor, pers. comm.) which represent two of the loci of the Hox-1 cluster in the mouse (Duboule et al., 1986). However, the observed recombination frequency of 18% (data not shown) between GR1 and the Hox-1 locus is beyond the efficient range of linkage detection in RI strains (Taylor, 1978). Similarly, comparison of the SDP of GR1 with the Ggc locus (Tulchin and Taylor, 1981) failed to reveal any tight linkage (data not shown). Ggc is known to be tightly linked with Hox-1 (Bucan et al., 1986).

Thus, the GR1 locus is expected to lie at some distance >12cM proximal to the Hox-1 cluster. The mapping data of Bucan et al. (1986) and Dean et al. (1987) are compatible with the inability to detect linkage between GR1 and Hox-1 in the BxD series of recombinant inbred strains (fig. 6.10). Given the proximal location of met on chromosome 6, it is possible that the range for the location of GR1 is limited by the centromere. The distance of met from the centromere is not known.

Figure 6.10. Location of GR1 locus on mouse chromosome 6. The linkage data presented in fig. 6.2 were used to locate the GR1 locus relative to *met* on mouse chromosome 6. The region in which GR1 is predicted to lie is designated by the arrows. The broken arrow represents the region predicted to lie within 12cM of the *Hox*-1 locus. The linkage data between *met* and other loci are derived from Bucan *et al.*, 1986 (right hand side) and Dean *et al.*, 1987 (left hand side). Distances are given in centiMorgans (cM). The distance between *met* and the centromere (black circle) is not known. *met*= *met* proto-oncogene, *Cpa*= carboxypeptidase A, *Tcrb*= T cell receptor β-chain, *Try*-1= trypsin, *Hox*-1= *Hox*-1 cluster, *IgK*= Immunoglobulin K, *raf*-1= proto-oncogene *raf*.


met
GRI
21± 7.8
Cpa
5
Try-1,
Tcrb
5
Hox-1
14±6.4
11
IgK
17±7.0
13
raf-1

More precise mapping of the GR1 locus to this region of mouse chromosome 6 could be carried out by determination of the SDP of Cpa, Tcrb, and Try-1 loci (fig. 6.10), and comparison with that of GR1. Additionally, segregation analyses of flanking markers relative to GR1 in interspecies backcross experiments could be used to give information concerning longer range linkage relationships between GR1 and other loci on mouse chromosome 6.

The major aim of mapping the GR1 locus to the mouse genome was to determine whether GR1 mapped close to any previously characterised developmental mutation. There are 7 such mutations on mouse chromosome 6. In order from the centromere these are: sightless (sig), hydrocephalic polydactyly (hpy), hypodactyly (hd), truncate (tc), postaxial hemimelia (px), microophthalmia (mi), and crooked (cd) (Green, 1981; Davisson and Roderick, 1986). Detailed mapping by 3 point cross analysis established the hd locus to lie between Tcrb and IgK (D'Hoostleare et al., 1985), close to the Hox-1 locus (Bucan et al., 1986). Given the absence of linkage between GR1 and Hox-1, only sig and hpy could be candidate loci for GR1. The phenotype of mice homozygous for the recessive mutation hpy includes preaxial polydactyly of fore and hind limbs, male sterility and hydrocephaly (Green, 1981). Presumed homozygotes for the semi-dominant mutation sig die at birth with pronounced hydrocephaly and often hind foot syndactyly (Green, 1981). Whilst such phenotypes may be compatible with the skeletal and neural expression of GR1 (section 5.2.4), no

defects in extraembryonic tissues of such mutants were reported. Given the results of RNA *in situ* hybridisation analyses (section 5.2.4), anomalies in such tissues might be expected if the GR1 locus was disrupted.

It should also be noted that *met* maps very close to the obese (*ob*) locus (Dean *et al.*, 1987). However, the phenotype of *ob* mice (Green, 1981) is not similar to that expected for a GR1-defective mouse, given the observed developmental expression of GR1 (section 5.2.4). Further mapping studies of GR1 relative to *sig*, *hpy* and *ob* would facilitate the construction of a detailed map of this region of mouse chromosome 6. It would also be interesting to determine if GR1 expression was abnormal in *sig* and *hpy* mutant embryos.

As mentioned earlier (section 6.1.3), comparative mapping between mouse and humans has revealed the conservation of a number of syntenic groups of genes. A number of loci on mouse chromosome 6 are found on human chromosome 7 (Searle *et al.*, 1987). *met*, *Cpa*, *Try*-1, and *Tcrb* loci on mouse chromosome 6 are found on the long arm of human chromosome 7 (Searle *et al.*, 1987). The distal limit to this group of loci is marked by *Hox*-1 and *Ggc* which map to the short arm of human chromosome 7 (Searle *et al.*, 1987). *IgK* and *raf*-1 map to human chromosomes 2 and 3 respectively (Malcolm *et al.*, 1982; Bonner *et al.*, 1984). In man, the proximal limit to the syntenic group on the long arm of chromosome 7 is marked by *Col*1A2 which maps to mouse chromosome 16 (Searle *et al.*, 1987).

In conjunction with further mapping studies in the mouse, it

would be of interest to determine the chromosomal location of any human homologue of GR1. This would provide additional information concerning the comparative organisation of conserved chromosomal elements on mouse chromosome 6 and human chromosome 7. Additional interest in mapping the human homologue of GR1 comes from the known tight linkage between *met* and the cystic fibrosis (CF) locus on human chromosome 7 (Dean *et al.*, 1985; White *et al.*, 1985). A large number of probes are available for this region of human chromosome 7 (Estivill *et al.*, 1987) and regions around the human *met* locus have been cloned (Michiels *et al.*, 1987). The availability of such reagents may facilitate the cloning of any human homologue of GR1. Moreover, the GR1 locus may provide a suitable marker for projects concerned with characterising a murine CF gene, with the aim of developing mouse models of cystic fibrosis (Dean *et al.*, 1987).

The tight linkage between GR1 and *met* in the mouse is also of interest with respect to the expression pattern of *met*. The human *met* oncogene was isolated from a chemically transformed human osteosarcoma derived cell line (Cooper *et al.*, 1984). In this cell line the gene is activated as a result of a translocation between chromosomes 1 and 6 (Park *et al.*, 1986). In the mouse, whilst *met* is not expressed in normal NIH3T3 fibroblasts, some transformed derivatives contain a 4-8 fold amplification of the *met* locus accompanied by increased levels of *met* RNA (Cooper *et al.*, 1986). Given the linkage between *met* and GR1, along with the restricted expression of GR1 in

murine fibroblasts (section 3.2.7), it would be of interest to determine the copy number and expression of GR1 in such transformed 3T3 cell lines.

Chapter 7. General discussion and future prospects.

From the data presented here, it is apparent that murine embryonal carcinoma cells can be utilised to isolate novel genes which exhibit regulated expression, not only on EC cell differentiation in vitro but also during normal mouse development. Whilst the approach of differential screening of cDNA libraries provided a pool of candidate genes of interest (Watson, 1986), the series of criteria outlined and tested in chapter 3 provided more stringent tests to establish which of these cDNA clones were of interest, within the context of the study of both transcriptional regulation of gene expression and gene function during mouse development.

Having found that GR1 met the various criteria for a "gene of interest", investigations of the expression pattern, structure and regulation of this novel gene were initiated. Whilst considerable progress has been made in such directions, it is clear that the data presented here represent only a preliminary analysis of this novel gene. However, they do provide the conceptual basis, and reagents, for subsequent studies. It is with the progress made and future directions for analysis of GR1 that this chapter is concerned.

Section 7.1. Regulation of GR1 expression in EC cells and during mouse development.

7.1.1. Transcriptional regulation of GR1 expression.

The clear down regulation of expression, both at the steady-state RNA and transcriptional levels, on chemically-induced differentiation of PCC3 cells (chapter 3), makes GR1 a suitable gene for the study of the mechanisms of transcriptional regulation which operate on low-copy cellular genes within murine embryo-derived stem cell lines.

Any analysis of the mechanism of transcriptional regulation demands the availability of the cis-regulatory elements of that gene. Thus, given that GR1 was isolated as a cDNA clone, it was necessary to isolate the corresponding genomic locus. However, the GR1 locus proved difficult to isolate from a number of cosmid and bacteriophage lambda libraries, and it was not until a bacteriophage lambda genomic library was constructed and screened on an McrB$^-$ host E.coli strain that the GR1 locus was recovered as the recombinant clone N28 (Raleigh et al., 1988; Chapter 4).

That N28 represents the GR1 locus has been established by restriction mapping, Southern blot and nucleotide sequence analyses. Mapping studies have established that N28 contains approximately 18kb of nucleotide sequence transcriptionally upstream of that represented by the cDNA clone pGR165. This genomic clone should prove useful in both regulatory and gene

function analyses of GR1.

Of prime importance regarding gene regulation studies is the determination of whether N28 contains the *cis*-acting regulatory elements of the GR1 transcription unit. Preliminary Southern blot analysis of N28, using radiolabelled RNA synthesised *in vitro* from PCC3 nuclei, suggests that this genomic clone does contain non-transcribed sequences (Reith and Stott, unpublished data). It should now be possible to determine the exon/intron organisation of this GR1 genomic clone by means of Northern blots and SI nuclease digestion experiments. The 5′ end of the GR1 transcription unit can be determined by use of primer extension analysis. The abundant levels of GR1 RNA detected in PCC3 cells ensure a rich, and easily accessible, source of GR1 RNA for such analyses.

Having determined the position of the transcriptional start site of the GR1 locus, sequences upstream of this site can be analysed to define the 5′ promoter/enhancer elements of the transcription unit.

Functional analysis could be carried out by means of transient assays for GR1 promoter activity in *in vitro* cell lines, such as the CAT assays used to characterise *cis*-elements of viral transcriptional regulatory elements in EC cells (section 1.3.2). Various restriction fragments of the putative GR1 5′ regulatory sequences would be fused with the *cat* structural gene and recombinant plasmids introduced into PCC3 EC and SV3T3 C138 cells by calcium phosphate-mediated DNA transfection (Gorman *et al.*, 1984). The SV3T3 C138 cell line

would serve as an easily accessible negative control, given the absence of GR1 RNA in this cell line (section 3.2). Any differences in transfection efficiency between cell lines could be monitored by the use of an appropriate control construct, such as RSV-cat (Gorman et al., 1985a,b), in parallel transfections.

In this way, it should be possible to define, at a crude level, the 5′ regulatory elements responsible for GR1 expression in PCC3 EC cells and/or repression in SV3T3 C138 cells. In addition, deletion analysis may distinguish elements required for cell type-specific expression from those necessary for basal level expression of GR1. The nucleotide sequence of the cis elements responsible for the regulated expression of GR1 in EC cells and fibroblasts can then be determined in order to establish any homology with DNA sequences known to bind specific trans-acting factors (Wingender, 1988).

Given the expression of GR1 in a number of tissues in the developing embryo (section 5.2), it should be possible to extend such CAT assay analyses of GR1 promoter activity to other cell types for which in vitro cell lines are available. For example, the cell line PYS-2 (Lehman et al., 1974) could be used to define GR1 cis-regulatory elements which mediate the high level expression of GR1 in parietal yolk sac during embryogenesis. Trophoblastoma (Nicolas et al., 1976) and trophectodermal carcinoma (Damjanov et al., 1985) cell lines could also be used to define cis-regulatory elements

responsible for the high level expression of GR1 in extraembryonic cell types. It may also be possible to utilise murine osteogenic (Sudo et al., 1983) and Swarm rat chondrosarcoma (Choi et al., 1971) cell lines as models for investigation of the transcriptional regulation of GR1 within developing bones.

Such analyses may indicate that the GR1 transcriptional regulatory region consists of one or more cell type-specific cis-regulatory elements. However, before CAT assay analysis of GR1 expression in such cell lines, preliminary Northern blot and transcriptional run-on analyses would be required in order to establish that the endogenous GR1 gene is expressed, and transcriptionally regulated, within these cell lines.

Should it be found that N28 does not contain the cis-regulatory sequences which mediate the observed expression pattern of this gene, it would be necessary to isolate additional GR1 genomic clones. The construction of the PCC3 genomic library in the bacteriophage lambda vector lambda fix (section 4.2.5) facilitates the generation of radiolabelled RNA probes specific for either end of a given genomic clone (fig. 4.10). Thus, a probe representing the 5′ end of N28 could be easily generated for the isolation of any additional clones representing regions of the GR1 locus transcriptionally upstream from those represented by N28.

Should 5′ non-transcribed sequences be found not to mediate appropriate expression of GR1, there may be cis-acting elements within introns and/or in 3′ untranslated sequences of

the GR1 locus. Given that N28 does not contain the 3′ end of the GR1 transcription unit, additional GR1 genomic clones would need to be isolated before analysis could proceed. The ease of detection of restriction fragment length polymorphisms at the 3′ end of the GR1 locus (section 6.2.6), and the absence of this region from the single GR1 genomic clone isolated, raises the possibility that the 3′ end of the GR1 locus represents a highly polymorphic region which may be unstable in E.coli. As such, this region of the GR1 locus may prove difficult to clone.

Any investigation of the roles of intron, 3′ untranslated or non-transcribed sequences in mediating GR1 expression would be best investigated by the use of GR1 mini-gene constructs in stable DNA transfection assays. It would be necessary to introduce a modification into the GR1 transcription unit in order to distinguish RNA derived from the experimental gene construct from that of the endogenous GR1 locus. As in CAT assays, deletion analyses would facilitate the characterisation of cis-regulatory elements. Moreover, such an approach, in conjunction with nuclear run-on analyses, could be used to determine cis-elements within the GR1 transcription unit which may mediate differential stability of GR1 RNA.

The cis regulatory elements defined by all of the above analyses can then be used in in vitro transcription systems and gel retardation assays, as used for model viral transcriptional regulatory elements (section 1.3.2), in order

to characterise and purify any *trans*-acting factors which may mediate the regulated expression of GR1 seen in *in vitro* cell lines.

7.1.2. Transgenesis as a route to the identification of GR1 *cis*-regulatory elements.

Whilst *in vitro* cell lines will prove essential to the analysis of the mechanism of transcriptional regulation of GR1 expression, not all sites of expression revealed by RNA *in situ* hybridisation analyses (section 5.2) can be represented by *in vitro* cell line models. For example, there is no easily manipulatable cell system which could be utilised to characterise the *cis*-regulatory elements responsible for the localised expression of GR1 in anterior, but not posterior sclerotome halves in the 12.5 days gestation mouse embryo. In addition, it will be necessary to confirm that any *cis*-regulatory elements defined by use of *in vitro* cell lines are also responsible for GR1 expression during normal mouse development, if regulatory studies are to be relevant.

As discussed earlier (section 1.4), transgenesis offers a means to analyse gene regulation and function during mouse embryogenesis. Initial transgenic experiments would be required in order to establish the extent of 5′ and 3′ non-transcribed sequences around the GR1 locus required to mediate all the sites of expression of this gene seen during mouse

development. The structural gene would require modification in order to distinguish GR1 transgene expression from that of the endogenous locus.

Subsequently, a number of deletion constructs could be introduced into the germline in order to define *cis*-elements required for appropriate temporal and spatial patterns of GR1 expression during embryogenesis. A rational basis for the design of such constructs could be provided from deletion analyses in *in vitro* cell lines (section 7.1.1).

It should be remembered that RNA *in situ* hybridisation analysis, as with Northern blots, reveals only differences in steady-state RNA levels between cell types. Thus, differential expression of GR1 seen during mouse development may be attributable to both transcriptional and post-transcriptional regulatory mechanisms. The inability to isolate large quantities of homogeneous cell types from the embryo precludes the use of transcriptional run-on analyses in isolated nuclei to establish transcriptional regulation as used for *in vitro* cell lines.

In this respect, the use of putative GR1 transcriptional regulatory sequences to express a structural marker gene, such as that for neomycin resistance, CAT or β-galactosidase, in transgenic mice would establish whether the differential expression observed at the steady-state RNA level was due to differential rates of transcription. Whilst such *cis*-regulatory elements will be essential for studies of the mechanism of transcriptional regulation, these DNA sequences

could also be used to redirect expression of a structural gene of interest to abnormal sites in transgenic mice in experiments designed to investigate gene function.

For example, the expression of many murine homeobox genes is spatially restricted to hind brain and thoracic regions along the anterior-posterior axis of the embryo (section 5.1.2). Expression of such murine homeobox genes under the transcriptional regulation of the GR1 promoter/enhancer elements in transgenic mice may result in homeobox expression along the entire rostro-caudal axis, as seen for the endogenous GR1 gene. Analysis of the phenotypic consequences of such deregulated expression would address the role of homeobox genes in mouse development. The cis-regulatory elements of GR1 could also be used to direct expression of the PAX-1 gene (Deutsch et al., 1988) to the anterior sclerotome half of somites. Such expression would be in addition to endogenous PAX-1 expression in the posterior sclerotome half, and may facilitate determination of the function of PAX-1 during development of the vertebral column.

The cis-regulatory sequences of the 1.6U gene(s) may also prove useful for such analyses of gene function. In contrast to GR1, the genomic clones of 1.6U represent a gene, the steady-state RNA levels of which are high in a variety of EC cells, differentiated cell lines and adult mouse tissues (section 3.2). From transcriptional run-on analyses in isolated nuclei of EC cells, and their differentiated derivatives, it is known that the 1.6U promoter(s) is

particularly active (A.R. Cooper, unpublished data). Characterisation of the cis-regulatory sequences of this gene, in a manner similar to that outlined for GR1, may provide transcriptional regulatory elements which can direct high constitutive expression of a given structural gene in a wide variety of embryonic and differentiated cell types. As such they could also be used to investigate the phenotypic consequences of deregulated expression of a given gene during mouse development (section 7.2.4).

Section 7.2. Function of GR1 in mouse development.

7.2.1. Expression pattern and sequence analyses of GR1.

From both Northern blot and RNA in situ hybridisation analyses (section 3.2; section 5.2), it is apparent that GR1 is expressed at relatively high levels at a number of discrete sites within the developing embryo, whereas expression is much more restricted in adult tissues.

Whilst a 6.8kb GR1 mRNA species was readily seen on Northern blot analysis, additional GR1 homologous RNA species were detected on long exposures of filters probed with the entire cDNA insert of pGR165 (section 3.2). Only some of these additional transcripts could be attributed to the poly-proline encoding repetitive element in pGR165, since probes lacking this sequence also hybridised to RNA species in PCC3 cells

smaller than the major 6.8kb GR1 mRNA (fig. 5.2).

Such observations may be the result of some degradation of the major GR1 mRNA species or may reflect transcripts derived from loci with weak homology to GR1, as represented by genomic clones N20.1 and N20.3 (section 4.2.6). The presence of consensus splice acceptor sites in the 3′ untranslated region of the GR1 transcription unit (section 6.3.1) raises the possibility that additional RNA species are derived by differential splicing of the primary transcript. Such possibilities should be borne in mind in the interpretation of RNA in situ hybridisation data.

The nature of any such differentially spliced or weakly homologous RNA species will be established by isolation and analysis of additional GR1 homologous cDNA clones. In this respect, the clones isolated from an 8.5 days gestation mouse embryo cDNA library (section 3.3.4) may prove useful. Restriction mapping, Northern blot, Southern blot and nucleotide sequence analyses will establish the relationship of these clones to existing GR1 cDNA and genomic clones.

The higher resolution of gene expression patterns determined by RNA in situ hybridisation, relative to Northern blot analyses, has been illustrated. However, it remains to be determined whether GR1 expression is maintained within specific cell populations in the adult mouse, as seen with the expression in the interstitial cells of adult testis.

For example, the low level of GR1 RNA detected in adult brain by Northern blot analysis may be attributable to maintained

expression of GR1 in the anterior lobe of the pituitary gland, as seen by RNA in situ hybridisation analysis of the mouse embryo. It is not known whether GR1 is expressed in adult cartilage or is restricted to the transient chondrocyte population of developing bones. Such expression will best be investigated by RNA in situ hybridisation analyses of adult mouse tissues. However, it is interesting to note that, whilst readily detectable in condensed mesenchyme around the bronchioles at 12.5 days gestation, GR1 expression was not detectable on Northern blot analysis of adult lung RNA. Thus, for lung at least, expression of GR1 is restricted to the embryonic stage of lung development and is not expressed in adult lung connective tissue.

From the expression pattern of GR1 during embryogenesis in membranes, connective tissues, chondrocytes, and mesenchymal cells associated with epithelia, it was tentatively concluded that GR1 may encode a component of the extracellular matrix (ECM) secreted by such cell types (section 5.3.3). Such a conclusion was supported by analysis of the amino acid sequence of the carboxy terminal region of the putative GR1 protein (section 6.3.2). Being rich in β-sheet secondary structures, it is likely that the GR1 protein is an extracellular non-enzymatic protein. However, on searching both nucleotide and amino acid data bases, the partial sequence of the GR1 transcription unit was found to have no homology with any previously characterised gene or protein (section 6.3). Thus, it is concluded that GR1 represents a

novel gene.

These data bases included a number of ECM components including sequences of the carboxy and 3′ untranslated regions of collagen types I, II, III, IV, V, IX, and X. Additionally, the nucleotide sequences of cDNA clones of chick type XII collagen have been determined (Gordon et al., 1987), but again no homologies between these sequences and that of GR1 were found. No sequence data of type VI collagen (Furthmayr et al., 1983) is available. Since the distribution pattern of this protein in the chick (Linsenmayer et al., 1986a,b) differs markedly from the RNA distribution of GR1 in the mouse, it is considered unlikely that GR1 is murine type VI collagen. However, the distribution of steady-state RNA levels need not be reflected in the distribution of a given gene product. Thus, in the absence of sequence data, it remains formally possible that GR1 is type VI collagen. A similar problem exists between any relationship of GR1 with type VII collagen, originally isolated from human amniotic membranes (Bentz et al., 1983), and type VIII collagen (Sage et al., 1983).

Collagens are defined by the presence of repetitive amino acid sequences of the sequence Gly-X-Y where X and Y often correspond to proline or hydroxyproline (see Bornstein and Sage, 1980; Cheah, 1985 for reviews). This sequence is responsible for the characteristic triple helical structure of collagens. Although no such sequences were found in 247 amino acids of the carboxy terminal region of the putative GR1 polypeptide, this may represent a propeptide or non-helical

domain of a collagen.

Formal proof that GR1 does not represent any known, or novel, collagen type will come from analysis of the GR1 genomic clone N28, or any additional clones containing as yet uncharacterised regions of the GR1 transcription unit. Southern blot analysis using a collagen triple helical probe would indicate any regions of the GR1 transcription unit containing such Gly-X-Y repetitive sequences. DNA sequence analysis of these regions would then confirm the open reading frame of such elements.

Another major component of extracellular matrices are proteoglycans (PG), of which a variety of core proteins are known (see Hassell et al., 1986 for review). The partial cDNA sequences of some PG core proteins, including the large cartilage-specific PG of the rat (Doege et al., 1986), a small PG core protein, PG19, isolated from a rat yolk sac tumour (Bourdon et al., 1985), and a human small PG (Krusius and Ruoslahti, 1986) were present in the data bases searched with the GR1 nucleotide and predicted amino acid sequences.

In addition, the entire nucleotide sequences of PG19 (Bourdon et al., 1986) and the large cartilage-specific PG (Doege et al., 1987), as well as sequence data of a second proteoglycan, PGII, isolated from a rat yolk sac tumour (Day et al., 1986) and a bovine cartilage PG (Oldberg et al., 1987) were found to have no homologies with GR1.

Like collagens, characteristic repetitive amino acid sequences have been found in proteoglycan core proteins. The dipeptide

sequences Gly-Ser or Gly-Thr are extensively repeated in a number of vertebrate PG core proteins (Robinson et al., 1978; Couchman et al., 1985; Seldin et al., 1985; Day et al., 1986; Bourdon et al., 1987a), as well as those of invertebrates (Jackson et al., 1986; Reddy et al., 1986), and are believed to represent attachment sites for the glycosaminoglycan side chains characteristic of proteoglycans (see Hassell et al., 1986 for review). Moreover, whilst the Gly-Ser coding region of PG19 is repeated numerous times in the rat genome, the carboxy terminal and 3′ untranslated regions of this gene are unique (Bourdon et al., 1987b). Thus, it is possible that GR1 represents a novel proteoglycan core protein which contains unique carboxy and 3′ untranslated domains. Southern blot and DNA sequence analyses of the GR1 genomic clone N28 will establish whether GR1 is a member of the proteoglycan core protein family of genes.

It should also be emphasised that the GR1 sequence reported here may represent a previously uncharacterised region of a known protein. The GR1 protein may have been purified and amino acid sequence of peptide fragments of regions other than the carboxy terminal domain may be present on current sequence data bases.

In this respect, there is an obvious need for further nucleotide sequence determination of the, as yet, uncharacterised 4.1kb of the GR1 transcript. Given that such analysis should concentrate on nucleotide and amino acid sequences located on the amino terminal side of those

determined in the present work, the GR1 genomic clone N28 will provide the necessary source of DNA for sequence determination. Such sequence analysis will complement the experiments outlined earlier (section 7.1.1) in the construction of an exon/intron map of the GR1 locus and isolation of cis-regulatory elements of this gene.

In addition to determining whether the sequence of any region of the GR1 transcription unit has been previously determined, or contains sequences characteristic of collagens or proteoglycans, analysis of the entire amino acid sequence of this gene product may reveal some structural similarities with other proteins, for example the Arg-Gly-Asp-Ser peptide shown to be the cell attachment site for fibronectin (Pierschbacher and Ruoslahti, 1984). Should no such primary sequence homologies be found, the predicted secondary and tertiary structure of the GR1 polypeptide may provide further clues as to the function of this gene product.

7.2.2. Antisera against GR1 protein(s).

Given the prediction from both expression and sequence analyses that GR1 may encode a novel component of the ECM, it is vital that antisera directed against this gene product be raised. Not only could such reagents be used to determine the cellular and tissue localisation of GR1 both in in vitro cell lines and mouse embryonic and adult tissues, but they would

also facilitate purification of GR1-encoded polypeptides from cell lines and tissues. Such antisera could be used to establish whether the high levels of GR1 RNA seen in many cell types are reflected in protein levels.

Given the high levels of GR1 RNA in extraembryonic tissues and chondrocytes of developing bones, it is possible that independent investigators have characterised the GR1 protein as a component of such tissues at the biochemical level, although no amino acid sequence data of such proteins may be available. The generation of GR1-specific antisera from knowledge of the predicted amino acid sequence presented here, will allow any putative relationships between GR1 and previously characterised basement membrane, cartilage or placental proteins, to be tested.

The need for GR1 antisera was anticipated, and attempts were made to generate such reagents, using the cDNA clone pGR165, before the nucleotide sequence and reading frame of this cDNA clone had been established (section 5.3.1). Although antisera were raised against the single stable β-galactosidase-GR1 fusion protein generated from the sense strand of a subclone of pGR165, subsequent sequence analysis indicated that the restriction fragment used consisted almost entirely of 3′ untranslated sequences of the GR1 transcription unit. This provides an explanation for the absence of any activity directed against antigens present in both PCC3 and SV3T3 C138 cells in the antiserum raised against this fusion protein (section 5.2.3).

Having determined the sequence of 247 amino acids of the putative GR1 polypeptide, it is now possible to design specific peptides and/or further fusion proteins which can be used in fresh efforts to generate GR1 antisera. Given the instability of the sense strand of the cDNA clone pGR165 in the pUR series of β-galactosidase fusion vectors (section 5.2.2), it may be prudent to pursue the use of synthetic peptides to raise GR1 antisera.

7.2.3. Poly-proline encoding repetitive element in the putative GR1 polypeptide.

From initial Southern blot analyses of mouse genomic DNA, it was established that the GR1 cDNA clone pGR165 contains a moderately repetitive element. This was localised to a 170 base pair restriction fragment (section 4.2.2) which contains a cytosine-guanine rich region, coding for a poly-proline sequence close to the carboxy terminus of the putative polypeptide (section 6.2). The unique properties of proline, and the unique conformation assumed by poly-proline peptides (section 6.3.2), makes this repetitive element worthy of further investigation with regard to the function of GR1.
The nature of the poly-proline helix is such as to facilitate the formation of multimeric protein complexes by means of hydrogen bonding between different polypeptide chains in a manner analogous to that seen with collagen triple helices.

Thus, the GR1 protein may exist as a multimeric protein *in vivo*. Alternatively, the poly-proline sequence may mediate interactions between GR1 and other proteins. The availability of GR1-specific antisera will facilitate the purification of the GR1 protein from cells and tissues and allow any such interactions to be investigated.

A number of proteins included in the July 1987 release of the EMBL data base contain proline-rich sequences, including salivary proline rich proteins, the c-*myc* proto-oncogene and type I and type II collagen chains. Such genes may account for a proportion of the genomic sequences seen to be homologous to the GR1 repetitive element in the mouse genome. Moreover, the genomic clones CAD18/20 and CAD22 (section 4.2.4) may represent the loci of such genes. Restriction mapping, Southern blot and nucleotide sequence determination would establish any relationship between these cosmid clones and previously characterised genes.

7.2.4. Analysis of GR1 function in transgenic mice.

It has recently been shown that mutation within the triple helical coding region of an α1(I) collagen gene can serve as a dominant negative mutation in transgenic mice, presumably by interfering with normal triple helix formation (Stacey *et al.*, 1988). By analogy, mutation within the poly-proline domain of GR1 may result in disruption of GR1 polymer formation, perhaps

with phenotypic consequences.

Such a hypothesis could be tested in transgenic mice. A functional GR1 gene, under the regulation of its own promoter/enhancer elements and bearing a mutation within the poly-proline encoding sequence, could be introduced into the fertilised egg by microinjection. Analysis of any phenotypic consequences of the expression of such a transgene would provide strong evidence in favour of a role for the poly-proline sequence in the function of GR1.

It will also be possible to study the effects of deregulated expression of a normal GR1 protein during mouse development. In this respect, the cis-regulatory elements of 1.6U could be used to direct GR1 expression to a wide variety of embryonic and adult tissues in transgenic mice. In addition, it may be possible to redirect GR1 expression in a more subtle manner in order to ask more specific questions concerning the role of this gene in particular aspects of mouse embryogenesis.

For example, it would be interesting to use the cis-regulatory elements of murine osteonectin (Mason et al., 1986) to express a GR1 transgene. RNA in situ hybridisation analyses have shown osteonectin to be expressed in osteoblasts of the periosteal and epiphysal surfaces of developing bones, but not within the chondrocyte population within embryonic long bones (Holland et al., 1987). Given that GR1 expression appears to be restricted to the chondrogenic stage of bone development, the use of an osteonectin-GR1 transgene would allow the phenotypic consequences of GR1 expression during the osteogenic stage of

bone development to be investigated.

From the known sites of c-*fos* expression in developing limbs (Dony and Gruss, 1987b), it may also be possible to direct GR1 expression to the perichondrium and mesodermal web tissue of developing limbs, in addition to the endogenous sites of GR1 expression. A similar approach could be used to re-direct GR1 expression to both posterior and anterior sclerotome halves by the use of the *cis*-regulatory elements of the *PAX*-1 gene (Deutsch *et al*., 1988) to express a GR1 transgene.

In all these cases, the phenotypic consequences of additional sites of GR1 expression on the developing embryo may prove informative in establishing a role for this gene in mouse development.

Functional gene inactivation experiments could also be carried out in order to investigate the function of GR1 during mouse embryogenesis. As discussed earlier, introduction of specific mutations into the GR1 coding sequence, such as those in the poly-proline domain, could prove useful with regard to analysis of GR1 function. However, given the expression pattern of GR1, the usefulness of targetted inactivation of the GR1 locus by homologous recombination is unclear.

Even if it were possible to disrupt the GR1 locus in ES cells and retain cell viability, investigation of the consequences of a complete loss of gene function in embryonic tissues may be precluded by the inability of extraembryonic tissues to function early in development. Whilst such a result would indicate that GR1 plays an essential role in the function of

extraembryonic tissues, it would not address the role of GR1 in, for example, gastrulation or bone development.

Given such problems, it might be thought that anti-sense RNA gene constructs could be utilised to specifically inactivate GR1 expression at particular sites during mouse embryogenesis. However, given that a 40:1 ratio of anti-sense to sense RNA was required in order to achieve a 50% reduction in HPRT protein in transgenic mice (Munir et al., 1987; C.T. Caskey, pers. comm.), it would seem unlikely that, at the present time, this could be achieved for GR1. Transcriptional regulatory elements are not currently available to facilitate the high levels of cell type-specific expression of anti-sense GR1 RNA required to inhibit GR1 function by this means in transgenic mice. However, the 1.6U cis-regulatory elements may prove useful in anti-sense experiments designed to inhibit translation of low abundance mRNA species.

7.2.5. Diversification of GR1 analyses into other experimental systems.

Given the limitations of the mouse embryo as a model system for studying the regulation of development of multicellular organisms (chapter 1), it may prove useful to investigate expression of any GR1 homologues in other vertebrate, as well as invertebrate, species amenable to experimental manipulation. As well as facilitating further analyses of the

function of this novel gene, conservation of GR1 across species would suggest that this gene may play an important role in fundamental biological processes.

In addition to the conservation of the homeobox domain and some proto-oncogenes between mammals and lower eukaryotes, Drosophila melanogaster has been found to contain genes homologous with vertebrate proteoglycans (Jackson et al., 1986; Reddy et al., 1986; Brower et al., 1987). Additionally, Drosophila Kc cells have been found to synthesise a number of proteins homologous to previously characterised vertebrate basement membrane proteins as well as a novel proteoglycan, papilin (Campbell et al., 1987).

The tight linkage observed between GR1 and the met proto-oncogene on mouse chromosome 6 justifies the investigation of the chromosomal position of any human homologue of GR1 (section 6.3.3). Allocation of a human homologue of GR1 to a specific region of the human genome would provide further information concerning the conservation of syntenic groups of loci between mouse chromosome 6 and human chromosome 7. Moreover, given the tight linkage of human met with the cystic fibrosis locus, this region of human chromosome 7 is currently subject to intense analyses, and a number of genomic clones of this region are available. Should GR1 and met show tight linkage in the human genome, it may be possible to utilise genomic clones of this region of human chromosome 7 in order to isolate the human GR1 gene. Additionally, establishment of linkage between GR1 and met in both mouse

and human would facilitate attempts to isolate and characterise a murine homologue of the CF locus.

In conclusion, it can be seen that the use of murine embryonal carcinoma cells has provided a route to the isolation of a novel, developmentally regulated, gene which can serve as a suitable substrate for the study of transcriptional regulation during the early stages of mouse embryogenesis. The availability of the cloned GR1 locus, along with the knowledge gained concerning the structure of the putative GR1 gene product and the expression pattern of this gene during mouse embryogenesis, will provide the framework upon which more detailed studies of gene regulation and function during mouse embryogenesis can be based.
Such efforts may play a small, but hopefully significant, role in the determination of the mechanisms by which a complexed multicellular organism, such as a mammal, is derived from the fertilised egg.

References.

Aaronson, S.A. and Todaro, G.J. 1968. Science 162: 1024

Adamson, E.D. 1987. Development 99: 449

Akam, M. 1987. Development 101: 1

Amundsen, S.K., Taylor, A.F., Chadhury, A.M. and Smith, G.R. 1986. Proc. Natl. Acad. Sci. USA 83: 5558

Andrews, P.W., Damjanov, I., Sunar, D., Banting, G., Carlin, C., Pracopoli, N. and Fogh, I. 1984. Lab. Invest. 50: 147

Angerer, L.M. and Angerer, R.C. 1981. Nuc. Acids Res. 9: 2819

Ann, D.K. and Carlson, D.M. 1985. J. Biol. Chem. 260: 15863

Ansorge, W. and Labeit, S. 1984. J. Biochem. Biophys. Meth. 10: 237

Appleyard, R.K. 1954. Genetics 39: 440

Atkinson, M.R., Deutscher, M.P., Kornberg, A., Russell, A.F. and Moffatt, J.G. 1969. Biochem. 8: 4897

Aviv, M. and Leder, P. 1972. Proc. Natl. Acad. Sci. USA 69: 1408

Awgulewitsch, A., Utset, M.F., Hart, C.P., McGinnis, W. and Ruddle, F.H. 1986. Nature 320: 328

Azen, E.A., Lush, I.E. and Taylor, B.A. 1986. Trends Genet. 2: 199

Bailey, D.W. 1971. Transplantation 11: 325

Barkai, U. and Sherman, M.I. 1987. J. Cell Biol. 104: 671

Barklis, E., Mulligan, R.C. and Jaenisch, R. 1986. Cell 47: 391

Barlow, D.P., Green, N.M., Kurkinen, M. and Hogan, B.L.M. 1984. EMBO J. 3: 2355

Baron, A., Featherstone, M.S., Hill, R.E., Hall, A., Gaillot, B. and Duboule, D. 1987. EMBO J. 6: 2977

Battula, N. and Loeb, L.A. 1975. J. Biol. Chem. 250: 4405

Beckwith, J.R. and Zipser, D. 1970. The lactose operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Belayew, A. and Tilghman, S.M. 1982. Mol. Cell. Biol. 2: 1427

Bender, W., Spierer, P. and Hogness, D.S. 1983. J. Mol. Biol. 168: 17

Benton, W.D. and Davis, R.W. 1977. Science 196: 180

Bentz, H., Morris, N.P., Murray, L.W., Sakai, L.Y., Hollister, D.W. and Burgeson, R.E. 1983. Proc. Natl. Acad. Sci. USA 80: 3168

Bernstine, E.G., Hooper, M.L., Grandchamp, S. and Ephrussi, B. 1973. Proc. Natl. Acad. Sci. USA 70: 3899

Biggin, M.D., Gibson, T.J. and Hong, G.F. 1983. Proc. Natl. Acad. Sci. USA 80: 3963

Birnhoim, H.C. and Doly, J. 1979. Nucleic Acids Research 7: 1513

Blumenthal, R. 1987. in Gene Amplification and Analysis. Vol.5 ed. Chirikjian, J.E. Elsevier, N.Y. pges 227-245.

Bonner, T., O'Brien, S.J., Nash, W.G., Rapp, U.R., Morton, C.C. and Leder, P. 1984. Science 223: 71

Boon, T., Buckingham, M.E., Dexter, T.L., Jakob, H. and Jacob, F. 1974. Ann. Microbiol. (Inst. Pasteur) 125B: 13

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. and Noll, M. 1986. Cell 47: 1033

Bornstein, P. and Sage, H. 1980. Ann. Rev. Biochem. 49: 957

Borrelli, E., Hen, R. and Chambon, P. 1984. Nature 312: 608

Bourdon, M.A., Oldberg, A., Pierschbacher, M. and Ruoslahti, E. 1985. Proc. Natl. Acad. Sci. USA 82: 1321

Bourdon, M.A., Shiga, M. and Ruoslahti, E. 1986. J. Biol. Chem. 261: 12534

Bourdon, M.A., Krusius, T., Campbell, S., Schwartz, N.B. and Ruoslahti, E. 1987a. Proc. Natl. Acad. Sci. USA 84: 3194

Bourdon, M.A., Shiga, M. and Ruoslahti, E. 1987b. Mol. Cell. Biol. 7: 33

Boyer, H.W. and Roulland-Dussoix, D. 1969. J. Mol. Biol. 41: 459

Bradley, A., Evans, M., Kaufman, M.H. and Robertson, E. 1984. Nature 309: 255

Bradley, A. and Robertson, E. 1986. Curr. Top. in Dev. Biol. 20: 357

Brahic, M. and Haase, A.T. 1978. Proc. Natl. Acad. Sci. USA 75: 6125

Brammar, W.J. 1982. in Genetic Engineering 3, ed. Williamson, R. A.P pages 53-81

Breathnach, R. and Chambon, P. 1981. Ann. Rev. Biochem. 50: 349

Brickell, P.M., Latchman, D.S., Murphy, D., Willison, K. and Rigby, P.W.J. 1983. Nature 306: 756

Brickell, P.M., Latchman, D.S., Murphy, D., Willison, K. and Rigby, P.W.J. 1985. Nature 316: 162

Brier, G., Bucan, M., Franke, U., Colberg-Poley, A.M. and Gruss, P. 1986. EMBO J. 5: 2209

Brigati, D.J., Myerson, D., Leary, J.J., Spalholz, B., Travis, S.Z., Fong, C.K.Y., Hsiung, G.D. and Ward, D.C. 1983. Virology 126: 32

Brinkley, L.L. and Morris-Wiman, J. 1984. Curr. Top. Dev. Biol. 19: 17

Brinster, R.L. 1972. in Growth, nutrition and metabolism of cells in culture. eds. Rothblat, G. and Cristafalo, V. Vol 2. New York, A.P.

Brinster, R.L. 1974. J. Exp. Med. 140: 1049

Bronner-Fraser, M. 1986. Dev. Biol. 115: 44

Broome, S. and Gilbert, W. 1978. Proc. Natl. Acad. Sci. USA 75: 2746

Brower, D.L., Piovant, M., Salatino, R., Brailey, J. and Hendrix, M.J.C. 1987. Dev. Biol. 119: 373

Brulet, P., Kaghad, M., Xu, Y-S., Croissant, O. and Jacob, F. 1983. Proc. Natl. Acad. Sci. USA 80: 5641

Brulet, P., Condamine, H. and Jacob, F. 1985. Proc. Natl. Acad. Sci. USA 82: 2054

Bucan, M., Yang-Feng, T., Colberg-Poley, A.M., Wolgemuth, D.J., Guenet, J-L., Francke, U. and Lehrach, H. 1986. EMBO J. 5: 2899

Burke, D.T., Carle, G.T. and Olsen, M.V. 1987. Science 236: 806

Burt, D.W., Reith, A.D. and Brammar, W.J. 1984. Nuc. Acids Res. 12: 8579

Callebaut, M. 1976. Experientia 32: 1337.

Campbell, A.G., Fessler, L.I., Salo, T. and Fessler, J.H. 1987. J. Biol. Chem. 262: 17605

Carbon, J., Clarke, L., Ilgen, C. and Ratzin, B. 1977. in Recombinant Molecules: Impact on Science and Society. eds. Beers, R.F. and Bassett, E. G. Raven Press, N.Y. pges 355-378

Carlon, N., Pizant, J. and Stahl, A. 1983. Anat. Embryol. 166: 399

Carter, A.T., Norton, J.D. and Avery, R.J. 1983. Nuc. Acids Res. 11: 6243

Chadhury, A.M. and Smith, G.R. 1984. Proc. Natl. Acad. Sci. USA 81: 7850

Cheah, K.S.E. 1985. Biochem. J. 229: 287

Chen, E. and Seeburg, P.H. 1985. DNA 4: 165

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. 1979. Biochemistry 18: 5294

Choi, H.U., Meyer, K. and Swarm, R. 1971. Proc. Natl. Acad. Sci. USA 68: 877

Chou, P.Y. and Fasman, G.D. 1974. Biochem. 13: 222

Clarke, L. and Carbon, J. 1976. Cell 9: 91

Cohen, F.E. and Sternberg, M.J.E. 1980. J. Mol. Biol. 137: 9

Cohen, F.E., Abarbanel, R.M., Kuntz, I.D. and Fletterick, R.J. 1983. Biochem. 22: 4894

Colberg-Poley, A.M., Voss, S.D., Chowdhury, K. and Gruss, P. 1985a. Nature 314: 713

Colberg-Poley, A.M., Voss, S.D., Chowdhury, K., Stewart, C.L., Wagner, E.F. and Gruss, P. 1985b. Cell 43: 39

Collins, F.S. and Weissman, S.M. 1984. Proc. Natl. Acad. Sci. USA 81: 6812

Collins, J. and Hohn, B. 1978. Proc. Natl. Acad. Sci. USA 75: 4242

Collins, J., Volckaert, G. and Nevers, P. 1982. Gene 19: 139

Cooper, C.S., Park, M., Blair, D.G., Tainsky, M.A., Huebner, K., Croce, C.M. and Vande Woude, G. 1984. Nature 311: 29

Cooper, C.S., Tempest, P.R., Beckman, M.P., Heldin, C-H. and Brookes, P. 1986. EMBO J. 5: 2623

Cosman, D. 1987. Immunol. Today 8: 16

Couchman, J.R., Woods, A., Hook, M. and Christner, J.E. 1985. J. Biol. Chem. 260: 13755

Coulson, A., Sulston, J., Brenner, S. and Karn, J. 1986. Proc. Natl. Acad. Sci. USA 83: 7821

Cowan, P.M. and McGavin, S. 1955. Nature 176: 501

Cowan, N.J., Dobner, P.R., Fuchs, E.V. and Cleveland, D.W. 1983. Mol. Cell. Biol. 3: 1738

Cox, K.H., DeLeon, D.V., Angerer, L.M. and Angerer, R.C. 1984. Dev. Biol. 101: 485

Cross, S.H. and Little, P.F.R. 1986. Gene 49: 9

Crouse, G.F. 1985. Gene 40: 151

Dale, R.B., McClure, B. and Houchins, J. 1985. Plasmid 13: 31

D'Amico-Martel, A. and Noden, D.M. 1983. Am. J. Anat. 166: 445

Damjanov, I., Damjanov, A. and Andrews, P.W. 1985. J. Embryol. exp. Morphol. 86: 125

Dautigny, A., Mattei, M-G., Morello, D., Alliel, P.M., Pham-Dinh, D., Amar, L., Arnaud, D., Simon, D., Mattei, J-F., Guenet, J-L., Jolles, P. and Avner, P. 1986. Nature 321: 867

Davisson, M.T. and Roderick, T.H. 1986. in Manipulating the mouse embryo. Hogan, B., Constantini, F. and Lacy, E. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.

Day, A.A., Ramis, C.I., Fisher, L.W., Gehron-Robey, P., Termine, J.D. and Young, M.F. 1986. Nuc. Acids Res. 14: 9861

Dean, M., Park, M., Le Beau, M.M., Robins, T.S., Diaz, M.O., Rowley, J.D., Blair, D.G. and Vande Woude, G.F. 1985. Nature 318: 385

Dean, M., Levine, R.A. and Campisi, J. 1986. Mol. Cell. Biol. 6: 518

Dean, M., Kozak, C., Robbins, J., Callahan, R., O'Brien, S. and Vande-Woude, G.F. 1987. Genomics 1: 167

Deschamps, J., DeLaaf, R., Joosen, L., Meijlink, F. and Destree, E. 1987. Proc. Natl. Acad. Sci. USA 84: 1304

Deutsch, U., Dressler, G.R. and Gruss, P. 1988. Cell, in press.

Dewey, M.J., Filler, R. and Mintz, B. 1978. Dev.Biol. 65: 230

D'Hoosteleare, L.A., Jouvin-Marche, E. and Huppi, K. 1985. Immunogenetics 22: 277

Diaz, M.O., Barsacchi-Pilone, G., Mahon, K.A. and Gall, J.G. 1981. Cell 24: 649

Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. 1983. Nuc. Acids Res. 11: 1475

Diwan, S.B. and Stevens, L.C. 1976. J. Natl. Canc. Inst. 57: 937

Doege, K., Fernandez, P., Hassell, J.R., Sasaki, M. and Yamada, Y. 1986. J. Biol. Chem. 261: 8108

Doege, K., Sasaki, M., Horigan, E., Hassell, J.R. and Yamada, Y. 1987. J. Biol. Chem. 262: 17757

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S. and Smithies, O. 1987. Nature 330: 576

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T. P., Bowden, D. W., Smith, D. R., Lander, E. S., Botstein, D., Akots, G., Rediker, S. K., Gravius, T., Brown, A. V., Rising, B. M., Parker, C., Powers, J. A., Watt, D. E., Kauffman, E. R., Bricker, A., Phipps, P., Muller-Kahle, H., Fulton, T. R., Ng, S., Schumm, J. W., Braman, J. C., Knowlton, R. G., Barker, D. F., Crooks, S. M., Lincoln, S. E., Daly, M. J. and Abrahamson, J. 1987. Cell 51, 319-337.

Donlon, T.A., Lalande, M., Wyman, A., Bruns, G. and Latt, S.A. 1986. Proc. Natl. Acad. Sci. USA 83: 4408

Dony, C., Kessel, M. and Gruss, P. 1985. Nature 317: 636

Dony, C. and Gruss, P. 1987a. EMBO J. 6: 2965

Dony, C. and Gruss, P. 1987b. Nature 328: 711

Doolittle, R.F. 1986. Of URFS and ORFS. University Science books, USA.

Doolittle, R.F. and Sapienza, C. 1980. Nature 284: 601

Duboule, D., Baron, A., Mahl, P. and Gaillot, B. 1986. EMBO J. 5: 1973

Ehrlich, M. and Wang, R.Y-H. 1981. Science 212: 1350

Engel, J. 1983. Biochem. J. 211: 303

Enquist, L.W. and Skalka, A. 1973. J. Mol. Biol. 75: 185

Erlich, H.A., Cohen, S.N. and McDevitt, H.O. 1978. Cell 13: 681

Estivill, X., Farrall, M., Scambler, P.J., Bell, G.M., Hawley, K.M.F., Lench, N.J., Bates, G.P., Kruyer, H.C., Frederick, P.A., Stanier, P., Watson, E.K., Williamson, R. and Wainwright, B.J. 1987. Nature 326: 840

Evans, M.J. and Kaufman, M.H. 1981. Nature 292: 154

Favalaro, J., Treisman, R. and Kamen, R. 1980. Methods in Enzymology 65: 718

Feinberg, A.P. and Vogelstein, B. 1984. Anal. Biochem. 137: 266

Fjose, A., McGinnis, W.J. and Gehring, W.J. 1985. Nature 313: 284

Flamant, F., Gurin, C.C. and Sorge, J.A. 1987. Mol. Cell. Biol. 7: 3548

Franz, T., Hilberg, F., Seliger, B., Stocking, C. and Ostertag, W. 1986. Proc. Natl. Acad. Sci. USA 83: 3292

Friedman, R.L., Manly, S.P., McMahon, M., Kerr, I.M. and Stark, G.R. 1984. Cell 38: 745

Friedman, S.A. and Hays, J.B. 1986. Gene 43: 255

Fujimura, F.K., Silbert, P.E., Eckhart, W. and Linney, E. 1981. J. Virol. 39: 306

Furthmayr, H., Wiedemann, H., Timpl, R., Odermatt, E. and Engel, J. 1983. Biochem. J. 211: 303

Gall, J.G. and Pardue, M.L. 1971. Methods in Enzymology 38: 470

Gardner, E. 1971. in The biochemistry and physiology of bone. second edition. Vol. 3. ed. Bourne, G.H.

Gardner, R.L. 1982. J. Embryol. exp. Morphol. 68: 175

Gardner, R.L. 1983. Int. Rev. Exp. Pathol. 24: 63

Gardner, R.L. and Rossant, J. 1979. J. Embryol. exp. Morph. 52: 141

Garnier, J., Osguthorpe, D.J. and Robson, B. 1978. J. Mol. Biol. 120: 97

Gaunt, S.J. 1987. Development 101: 51

Gaunt, S.J., Miller, J.R., Powell, D.J. and Duboule, D. 1986. Nature 324: 662

Gehring, W.J. 1985. Cell 40: 3

Gerard, G.F. 1986. Focus 8(3): 12

Gerhard, D.S., Kawasaki, E.S., Bancroft, F.C., Szabo, P. 1981. Proc. Natl. Acad. Sci. USA 78: 3755

Giguere, V., Ong, E.S., Segui, P. and Evans, R.M. 1987. Nature 330: 624

Godbout, R., Ingram, R. and Tilghman, S.M. 1986. Mol. Cell. Biol. 6: 477

Gooi, H.C., Feizi, T., Kapadia, A., Knowles, B.B., Solter, D. and Evans, M.J. 1981. Nature 292: 156

Gopinathan, K.P., Weymouth, L.A., Kunkel, T.A. and Loeb, L.A. 1979. Nature 278: 857

Gordon, J.W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A. and Ruddle, F.H. 1980. Proc. Natl. Acad. Sci. USA 77: 7380

Gordon, J.W. 1986. Mol. Cell. Biol. 6: 2158

Gordon, M.K., Gerecke, D.R. and Olsen, B.R. 1987. Proc. Natl. Acad. Sci. USA 84: 6040

Gorman, C.L., Moffat, L. and Howard, B. 1982. Mol. Cell. Biol. 2: 1044

Gorman, C.L., Lane, D.P. and Rigby, P.W.J. 1984. Proc. R. Soc. Lond. B 307: 343

Gorman, C.M., Lane, D.P., Watson, C.J. and Rigby, P.W.J. 1985a. Cold Spring Harb. Symp. Quant. Biol. 50: 701

Gorman, C.L., Rigby, P.W.J. and Lane, D.P. 1985b. Cell 42: 519

Gossler, A., Doetschman, T., Korn, R., Serfling, E. and Kemler, R. 1986. Proc. Natl. Acad. Sci. USA 83: 9065

Gray, M.R., Colot, H.V., Guarente, L. and Rosbash, M. 1982. Proc. Natl. Acad. Sci. USA 79: 6598

Green, M.C. 1981. Genetic variants and strains of the laboratory mouse. Fischer. Stuttgart; New York.

Grosveld, F.G., Lund, T., Murray, E.J., Mellor, A.J., Dahl, H.H.M. and Flavell, R.A. 1982. Nuc. Acids Res. 10: 6715

Grosveld, F., Blom van Assendelft, G., Greaves, D.R. and Kollias, G. 1987. Cell 51: 975

Gruneberg, H. 1963. The pathology of development. Blackwell, Oxford.

Grunstein, M. and Hogness, D. 1975. Proc. Natl. Acad. Sci. USA 72: 3961

Guarente, L. and Ptashne, M. 1981. Proc. Natl. Acad. Sci. USA 78: 2199

Haber, D.A., Beverly, S.M., Kiely, M.L. and Schmike, R.T. 1981. J. Biol. Chem. 256: 9501

Hafen, E., Levine, M., Garber, R.L. and Gehring, W.J. 1983. EMBO J. 2: 617

Hafen, E., Levine, M. and Gehring, W.J. 1984a. Nature 307: 287

Hafen, E., Kuroiwa, A. and Gehring, W.J. 1984b. Cell 37: 833

Hall, B.K. 1982. J. Embryol. exp. Morph. 68: 127

Hall, B.K. 1986. in Somites in developing embryos. eds. Bellairs, R., Ede, D.A. and Lash, J.W. Plenum Press.

Hammer, R.E., Krumlauf, R., Camper, S.A., Brinster, R.L. and Tilghman, S.M. 1987. Science 235: 53

Hanahan, D., Lane, D., Lipsich, L., Wigler, M. and Botchan, M. 1980. Cell 21: 127

Harbers, K., Kuehn, M., Delius, H. and Jaenisch, R. 1984. Proc. Natl. Acad. Sci. USA 81: 1504

Harding, J.D., MacDonald, R.J., Przybyla, A.E., Chirgwin, J.M., Pictet, R.L. and Rutter, W.J. 1977. J. Biol. Chem. 252: 7391

Harrison, P.R., Conkie, D., Paul, J. and Jones, K. 1973. FEBS Letts. 32: 109

Harrison, P.R., Conkie, D., Affara, N. and Paul, J. 1974. J. Cell Biol. 63: 402

Hartung, S., Jaenisch, R. and Breindl, M. 1986. Nature 320: 365

Hassell, J.R., Kimura, J.H. and Hascall, V.C. 1986. Ann. Rev. Biochem. 55: 539

Heikkila, R., Schawb, G., Wickstrom, E., Loke, S.L., Pluznik, D.H., Watt, R. and Neckers, L.M. 1987. Nature 328: 445

Heinegard, D. and Paulsson, M. 1987. Methods in Enzymology 145: 336

Hen, R., Borrelli, E., Fromental, C., Sassone-Corsi, P. and Chambon, P. 1986. Nature 321: 249

Herbommel, P., Bourachot, B. and Yaniv, M. 1984. Cell 39: 653

Herskowitz, I. 1987. Nature 329: 219

Hilberg, F., Stocking, C., Ostertag, W. and Grez, M. 1987. Proc. Natl. Acad. Sci. USA 84: 5232

Hogan, B.L.M., Taylor, A. and Adamson, E. 1981. Nature 291: 235

Hogan, B.L.M., Barlow, D.P. and Tilly, R. 1983. Canc. Surv. 2: 115

Hogan, B.L.M., Holland, P. and Schofield, P. 1985. Trends Genet. 1: 67

Hogan, B.L.M., Consantini, F. and Lacy, E. 1986. Manipulating the mouse embryo. Cold Spring Harbor Laboratory. Cold Spring Harbor, N.Y.

Hojman-Montes de Oca, F., Dianoux, L., Peries, J. and Emanoil-Ravicovitch, R. 1983. J. Virol. 46: 307

Holland, P.W.H., Harper, S.J., McVey, J.H. and Hogan, B.L.M. 1987. J. Cell Biol. 105: 473

Hooper, M., Hardy, K., Handyside, A., Hunter, S. and Monk, M. 1987. Nature 326: 292

Hopp, T.P. and Woods, K.R. 1981. Proc. Natl. Acad. Sci. USA 78: 3824

Howe, C.C. and Overton, G.C. 1986. Mol. Cell. Biol. 6: 150

Huszar, D., Balling, R., Kothary, R., Magli, M.C., Hozumi, N., Rossant, J. and Bernstein, A. 1985. Proc. Natl. Acad. Sci. USA 82: 8587

Huynh, T.V., Young, R.A. and Davis, R.W. 1985. in DNA cloning vol. I, ed. Glover, D.M. IRL Press.

Hyafil, F., Babinet, C., Huet, C. and Jacob, F. 1983. Cold Spring Harbor Conf. Cell Prolif. 10: 197

Imperiale, M.J., Kao, H-T., Feldman, L.T., Nevins, J.R. and Strickland, S. 1984. Mol. Cell. Biol. 4: 867

Ingham, P.W., Howard, K.R. and Ish-Horowicz, D. 1985. Nature 318: 439

Isemura, S., Saitoh, E. and Sanada, K. 1979. J. Biochem. 86: 79

Israel, D.I. and Whitelock, J.P. 1984. J. Biol. Chem. 259: 5400

Ito, K. and Takeuchi, T. 1984. J. Embryol. exp. Morphol. 84: 49

Izant, J.G. and Weintraub, H. 1984. Cell 36: 1007

Jackson, F.R., Bargiello, T.A., Yun, S-H., Young, M.W. 1986. Nature 320: 185

Jacob, F. 1979. Curr. Top. Dev. Biol. 13: 117

Jaenisch, R. 1976. Proc. Natl. Acad. Sci. USA 73: 1260

Jaenisch, R., Harbers, K., Schnieke, A., Lohler, J., Chumakov, I., Jahner, D., Grotkopp, D. and Hoffmann, E. 1983. Cell 32: 209

Jaenisch, R. and Jahner, D. 1984. Biochem. Biophys. Acta. 782: 1

Jahner, D., Haase, D., Mulligan, R. and Jaenisch, R. 1985. Proc. Natl. Acad. Sci. USA 82: 6927

Jakob, H., Boon, T., Gaillard, J., Nicolas, J-F. and Jacob, F. 1973. Ann. Microb.(Inst. Pasteur) 124B: 269

Jakobovits, A., Shackleford, G.M., Varmus, H.E. and Martin, G.R. 1986. Proc. Natl. Acad. Sci. 83: 7806

Jetten, A.M and Jetten, M.E.R. 1979. Nature 278: 182

Jetten, A.M., Jetten, M.E.R. and Sherman, M.I. 1979. Exp. Cell Res. 124: 381

John, H.A., Patrinou-Georgoulas, M. and Jones, K.W. 1977. Cell 12: 501

Johnson, M.L. 1933. Am. J. Anat. 52: 241

Johnson, D.R. 1986. The genetics of the skeleton. Clarendon Press, Oxford.

Johnson, M.H. and Maro, B. 1986. in Experimental approaches to mammalian embryonic development. eds. Rossant, J. and Pedersen, R.A. CUP.

Johnson, M.L. 1933. Am. J. Anat. 52: 241

Jones, T.R. and Cole, M.D. 1987. Mol. Cell. Biol. 7: 4513

Jones-Villeneuve, E.M.V., Rudnicki, M.A., Harris, J.F. and McBurney, M.W. 1983. Mol. Cell. Biol. 3: 2271

Joyner, A.L. and Bernstein, A. 1983. Mol. Cell. Biol. 3: 2180

Joyner, A.L., Kornberg, T., Coleman, K.G., Cox, D.R. and Martin, G.R. 1985. Cell 43: 29

Joyner, A.L. and Martin, G.R. 1987. Genes Dev. 1: 29

Kabsch, W. and Sander, C. 1983. FEBS Lett. 155: 179

Kabsch, W. and Sander, C. 1984. Proc. Natl. Acad. Sci. USA 81: 1075

Kaghad, M., Maillet, L. and Brulet, P. 1985. EMBO J. 4: 2911

Karn, J., Matthes, H.W.D., Gait, M.J. and Brenner, S. 1984. Gene 32: 216

Kelly, F. and Condamine, H. 1982. Biochem. Biophys. Acta 651: 105

Kelly, S.J. 1977. J. Exp. Zool. 200: 365

Kemp, D.J. and Cowman, A.F. 1981. Proc. Natl. Acad. Sci. USA 78: 4520

Keshet, E. and Itin, A. 1982. J. Virol. 43: 50

Keynes, R.J. and Stern, C.D. 1984. Nature 310: 786

Keynes, R.J. and Stern, C.D. 1985. Trends Neurosci. 8: 220

Keynes, R.J. and Stern, C.D. 1986. in Somites in developing embryos. eds. Bellairs, R., Ede, D.A. and Lash, J.W. Plenum Press.

Knecht, D.A. and Loomis, W.F. 1987. Science 236: 1081

Koenen, M., Ruther, U. and Muller-Hill, B. 1982. EMBO J. 1: 509

Koenen, M., Griesser, H-W., Muller-Hill, B. 1985. in DNA Cloning I. ed. Glover, D.M. IRL Press, Oxford. pges 89-100

Kondoh, H., Takahashi, Y. and Okada, T.S. 1984. EMBO J. 3: 2009

Konecki, D.S., Brennand, J., Fuscoe, J.C., Caskey, C.T. and Chinault, A.C. 1982. Nuc. Acids Res. 10: 6763

Kornberg, A. 1982. DNA replication (suppl.) S26. Freeman, San Francisco.

Korneluk, R.G., Quan, F. and Gravel, R.A. 1985. Gene 40: 317

Kramerov, D.A., Tillib, S.V., Ryskov, A.P. and Georgiev, G.P. 1985. Nuc. Acids Res. 13: 6423

Kratochwil, K., Dziadek, M., Lohler, J., Harbers, K. and Jaenisch, R. 1986. Dev. Biol. 117: 596

Krumlauf, R., Hammer, R.E., Tilghman, S.M. and Brinster, R.L. 1985. Mol. Cell. Biol. 5: 1639

Krumlauf, R., Holland, P.W.H., McVey, J.H. and Hogan, B.L.M. 1987. Development 99: 603

Krusius, T. and Ruoslahti, E. 1986. Proc. Natl. Acad. Sci. USA 83: 7683

Kuehn, M.R., Bradley, A., Robertson, E.J. and Evans, M.J. 1987. Nature 326: 295

Kurikinen, M., Barlow, D.P., Helfman, D.M., Williams, J.G. and Hogan, B.L.M. 1983. Nuc. Acids Res. 11: 6199

Kyte, J. and Doolittle, R.F. 1982. J. Mol. Biol. 157: 105

Laemmli, U.K. 1970. Nature 227: 680

La Femina, R. and Hayward, G.S. 1986. J. Virol. 58: 434

Lalley, P.A., Francke, U. and Minna, J.D. 1978. Proc. Natl. Acad. Sci. USA 75: 2382

Lam, S.T., Stahl, M.M., McMilin, K.D. and Stahl, F.W. 1974. Genetics 77: 425

Lash, J.W. and Yamada, K.M. 1986. in Somites in developing embryos. eds. Bellairs, R., Ede, D.A. and Lash, J.W. Plenum Press.

Laskey, R.A. and Mills, R.D. 1977. FEBS Letters 82: 314

La Thangue, N.B. and Rigby, P.W.J. 1987. Cell 49: 507

Lauer, J., Shen, C-K. J. and Maniatis, T. 1980. Cell 20: 119

Leach, D.R.F. and Stahl, F. 1983. Nature 305: 448

Le Beau, M.M., Westbrook, C.A., Diaz, M.O., Larson, R.A., Rowley, J.D., Gasson. J.C., Golde, D.W. and Sherr, C.J. 1986. Science 231: 984

Le Beau, M.M., Epstein, N.D., O'Brien, S.J., Neinhuis, A.W., Yang, Y-C., Clark, S.C. and Rowley, J.D. 1987. Proc. Natl. Acad. Sci. USA 84: 5913

Le Douarin, N.M. 1982. The Neural Crest. C.U.P.

Lehman, J.M., Speers, W.C., Swartzendruber, D.E. and Pierce, G.B. 1974. J. Cell. Physiol. 84: 13

Lehrach, H., Diamond, D., Wazney, J.M. and Boedtker, M. 1977. Biochemistry 16: 4743

Le Lievre, C.S. 1978. J. Embryol. exp. Morphol. 47: 17

Le Lievre, C.S. and Le Douarin, N.M. 1975. J. Embryol. exp. Morphol. 34: 125

Le Mouellic, H., Condamine, H. and Brulet, P. 1988. Genes Dev. 2: 125

Levine, R.A., LaRosa, G.J. and Gudas, L.J. 1984. Mol. Cell. Biol. 4: 2142

Lim, V.I. 1974. J. Mol. Biol. 88: 873

Lim, T.M., Lunn, E.R., Keynes, R.J. and Stern, C.D. 1987. Development 100: 525

Linnenbach, A., Huebner, K. and Croce, C.M. 1980. Proc. Natl. Acad. Sci. USA 77: 4875

Linney, E. and Donerly, S. 1983. Cell 35: 693

Linney, E., Davis, B., Overhauser, J., Chao, E. and Fan, H. 1984. Nature 308: 470

Linsenmayer, T.F., Mentzer, A., Irwin, M.H., Waldrep, N.K. and Mayne, R. 1986a. Exp. Cell Res. 165: 518

Linsenmayer, T.F., Bruns, R.R., Mentzer, A. and Mayne, R. 1986b. Dev. Biol. 118: 425

Liu, L-S. and Lark, K.G. 1982. Mol. Gen. Genet. 188: 27

Loenen, W.A.M. and Brammar, W.J. 1980. Gene 10: 249

Loenen, W.A.M. and Blattner, F.R. 1983. Gene 26: 171

Loh, T.P., Sievert, L.L. and Scott, R.W. 1987. Mol. Cell. Biol. 7: 3775

Lohler, J., Timpl, R. and Jaenisch, R. 1984. Cell 38: 597

Lovell-Badge, R.H. and Evans, M.J. 1980. J. Embryol. Exp. Morphol. 59: 187

Magnusson, T. 1986. in Experimental approaches to mammalian embryonic development. eds. Rossant, J. and Pedersen, R.A. CUP.

Malcolm, S., Barton, P., Murphy, C., Ferguson-Smith, M.A., Bentley, D.L. and Rabbitts, T.H. 1982. Proc. Natl. Acad. Sci. USA 79: 4957

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis. 1978. Cell 15: 685

Maniatis, T., Fritsch, E.F. and Sambrook, J. 1982. Molecular Cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., USA.

Manley, J.L., Fire, A., Cano, A., Sharp,P.A., Gefter, M.L. 1980. Proc. Natl. Acad. Sci. USA 77: 3855

Mann, R., Mulligan, R.C. and Baltimore, D. 1983. Cell 33: 153

Mark, W.H., Signorelli, K. and Lacy, E. 1985. Cold Spring Harb. Symp. Quant. Biol. 50: 453

Martin, G.R. 1981. Proc. Natl. Acad. Sci. USA 78: 7634

Martin, G.R. and Evans, M.J. 1975. Proc. Natl. Acad. Sci. USA 72: 3585

Martin, G.R., Boncinelli, E., Duboule, D., Gruss, P., Jackson, I., Krumlauf, R., Lonai, P., McGinnis, W., Ruddle, F. and Wolgemuth, D. 1987. Nature 325: 21

Mason, I.J., Murphy, D, and Hogan, B.L.M. 1985. Differentiation 30: 76

Mason, I.J., Taylor, A., Williams, J.G., Sage, H. and Hogan, B.L.M. 1986. EMBO J. 5: 1465

Maxam, A.M. and Gilbert, W. 1977. Proc. Natl. Acad. Sci. USA 74: 560

May, M.S. and Hattman, S. 1975. J. Bact. 123: 768

McBurney, M.W. and Rogers, B.J. 1982. Dev. Biol. 89: 503

McBurney, M.W., Jones-Villeneuve, E.M.V., Edwards, M.K.S. and Anderson, P.J. 1982. Nature 299: 165

McCue, P.A., Matthaei, K.I., Taketo, M. and Sherman, M.I. 1983. Dev. Biol. 96: 416

McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. and Gehring, W.J. 1984a. Nature 308: 428

McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A. and Gehring, W.J. 1984b. Cell 37: 403

McKnight, G.S. and Palmiter, R.D. 1979. J. Biol. Chem. 254: 9050

McKusick, V.A. and Ruddle, F.H. 1977. Science 196: 390

Meier, S. and Tam, P.P.L. 1982. Differentiation 21: 95

Meijer, D., Hermans, A., Von Lindern, M., Van Agthoven, T., de Klein, A., Mackenbach, P., Grootgoed, A., Talarico, D., Della Valle, G. and Grosveld, G. 1987. EMBO J. 6: 4041

Meijlink, F., Curran, T., Miller, D.A. and Verma, I.M. 1985. Proc. Natl. Acad. Sci. USA 82: 4987

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. 1984. Nucleic Acids Research 12: 7035

Messing, J. 1983. Methods in Enzymology 101: 20

Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. 1977. Proc. Natl. Acad. Sci. USA 74: 3642

Messing, J., Crea, R. and Seeburg, P.H. 1981. Nuc. Acids Res. 9: 309

Messing, J. and Vieira, J. 1982. Gene 19: 269

Michiels, F., Burmeister, M. and Lehrach, H. 1987. Science 236: 1305

Mills, D.R. and Kramer, F.R. 1979. Proc. Natl. Acad. Sci. USA 76: 2232

Mintz, B. and Ilmensee, K. 1975. Proc. Natl. Acad. Sci. USA 72: 3585

Mizusawa, S., Nishimura, S. and Seela, F. 1986. Nuc. Acids Res. 14: 1319

Mole, S.E. and Lane, D.P. 1985. J.Virol. 54: 703

Molineaux, S.M., Engh, H., De Ferra, F., Hudson, L. and Lazzarini, R.A. 1986. Proc. Natl. Acad. Sci. USA 83: 7542

Morange, M., Diu, A., Bensaude, O. and Babinet, C. 1984. Mol. Cell. Biol. 4: 730

Morgan, R.A., Christy, R.J. and Huang, R.C.C. 1988. Development 102: 23

Muller, R. 1986. Biochem. Biophys. Acta 823: 207

Mullins, J.J., Burt, D.W., Windass, J., McTurk, P., George, H. and Brammar, W.J. 1982. EMBO J. 1: 1461

Mullis, K.B. and Faloona, F.A. 1987. Methods in Enzymology 155: 335

Munir, M.I., Stout, J.T. and Caskey, C.T. 1987. Ann. Soc. Human Gen. (Abstr.)

Murphy, D., Brickell, P.M., Latchman, D.S., Willison, K. and Rigby, P.W.J. 1983. Cell 35: 865

Murray, N.E., Brammar, W.J. and Murray, K. 1977. Mol. Gen. Genet. 150: 53

Nadeau, J.H. and Taylor, B.A. 1984. Proc. Natl. Acad. Sci. USA 81: 814

Nader, W.F., Edlind, T.D., Huettermann, A. and Sauer, H.W. 1985. Proc. Natl. Acad. Sci. USA 82: 2698

Nicolas, J-F., Dubois, P., Jakob, H., Gaillard, J. and Jacob, F. 1975. Ann. Microb.(Inst. Pasteur) 126A: 3

Nicolas, J.F., Avner, P., Gaillard, J., Guenet, J.L., Jakob, H. and Jacob, F. 1976. Canc. Res. 36: 4224

Nikaido, T., Nakai, S. and Honjo, T. 1981. Nature 292: 845

Nishikawa, K. 1983. Bioch. Biophys. Acta 748: 285

Nishikawa, K. and Ooi, T. 1982. J. Biochem. 91: 1821

Nishikawa, K., Kubota, Y. and Ooi, T. 1983. J. Biochem. 94: 997

Nishikura, K. and Murray, J.M. 1987. Mol. Cell. Biol. 7: 639

Noden, D.M. 1983a. Dev. Biol. 96: 144

Noden, D.M. 1983b. Am. J. Anat. 168: 257

Norrander, J., Kempe, T. and Messing, J. 1983. Gene 26: 101

Oldberg, A., Antonsson, P. and Heinegard, D. 1987. Biochem. J. 243: 255

Olsen, M.V., Dutchik, J.E., Graham, M.Y., Brodeur, G.M., Helms, C., Frank, M., MacCollin, M., Scheinman, R. and Frank, T. 1986. Proc. Natl. Acad. Sci. USA 83: 7826

Olsson. A., Moks, T., Muhlen, M. and Gaal, A.B. 1984. J. Biochem. Biophys. Meth. 10: 83

Orgel, L.E. and Crick, F.H.C. 1980. Nature 284: 604

Orr-Weaver, T.L., Szostak, J.W. and Rothstein, R.J. 1981. Proc. Natl. Acad. Sci. USA 78: 6354

Overbeek, P.A., Lai, S-P., Van Quill, K.R. and Westphal, H. 1986. Science 231: 1574

Palmiter, R.D., Wilkie, T.M., Chen, H.Y. and Brinster, R.L. 1984. Cell 36: 869

Palmiter, R.D. and Brinster, R.L. 1986. Ann. Rev. Genet. 20: 465

Papaioannou, V.E., McBurney, M.W., Gardner, R.L. and Evans, M.J. 1975. Nature 258: 70

Park, M., Dean, M., Cooper, C.S., Schmidt, M., O'Brien, S.J., Blair, D.G. and Vande Woude, G.F. 1986. Cell 45: 895

Paulin, D., Perreau, J., Jakob, H., Jacob, F. and Yaniv, M. 1979. Proc. Natl. Acad. Sci. USA 76: 1891

Pauling, L., Corey, R.B. and Branson, H.R. 1951. Proc. Natl. Acad. Sci. USA 37: 205

Pellicer, A., Wigler, M., Axel, R. and Silverstein, S. 1978. Cell 14: 133

Perlumter, R.M., Ram, D. and Hood, L. 1984. in Genes and Cancer. UCLA Symposium on Molecular and Cellular Biology, new series. eds. Bishop, J.M., Rowley, J.D. and Greaves, M. (Liss, N.Y), pages 489-499

Petkovitch, M., Brand, N.J., Krust, A. and Chambon, P. 1987. Nature 330: 444

Pettenati, M.J., Le Beau, M.M., Lemons, R.S., Shima, E.A., Kawasaki, E.S., Larson, R.A., Sherr, C.J., Diaz, M.O. and Rowley, J.D. 1987. Proc. Natl. Acad. Sci. USA 84: 2970

Pierschbacher, M.D. and Ruoslahti, E. 1984. Nature 309: 30

Piko, L. and Clegg, K.B. 1982. Dev. Biol. 89: 362

Piko, L., Hammons, M.D. and Taylor, K.D. 1984. Proc. Natl. Acad. Sci. USA 81: 488

Pratt, D. 1969. Ann. Rev. Genetics 3: 343

Raleigh, E.A. and Wilson, G. 1986. Proc. Natl. Acad. USA 83: 9070

Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., Rigby, P.W.J., Elhai, J. and Hanahan, D. 1988. Nuc. Acids Res. 16: 1563

Rao, S.T. and Rossman, M.G. 1973. J. Mol. Biol. 76: 241

Readhead, C., Popko, B., Takahashi, N., Shine, H.D., Saavedra, R.A., Sidman, R.L. and Hood, L. 1987. Cell 48: 703

Reddy, P., Jacquier, A.C., Abovich, N., Petersen, G. and Rosbash, M. 1986. Cell 46: 53

Revel, H.R. 1967. Virology 31: 688

Richardson, J.S. 1981. Adv. Prot. Chem. 34: 167

Rickmann, M., Fawcett, J.W. and Keynes, R.J. 1985. J. Embryol. exp. Morphol. 90: 437

Rigby, P.W.J., Chia, W., Clayton, C.E. and Lovett, M. 1980. Proc. R. Soc. Lond. B. 210: 437

Robert, B., Barton, P., Minty, A., Daubas, P., Weydert, A., Bonhomme, F., Catalan, J., Chazottes, D., Guenet, J-L. and Buckingham, M. 1985. Nature 314: 181

Robinson, H.C., Horner, A.A., Hook, M., Ogren, S. and Lindahl, U. 1978. J. Biol. Chem. 253: 6687

Rodemer, E.S., Ihmer, A. and Wartenberg, H. 1986. J. Embryol. exp. Morphol. 98: 269

Rose, M., Casadaban, M.J. and Botstein, D. 1981. Proc. Natl. Acad. Sci. USA 78: 2460

Rosenberg, U.B., Preiss, A., Seifert, E., Jackle, H. and Knipple, D.C. 1985. Nature 313: 703

Rosenthal, A., Wright, S., Cedar, H., Flavell, R. and Grosveld, F. 1984. Nature 310: 415

Rossant, J. and McBurney, M.W. 1982. J. Embryol. Exp. Morphol. 70: 99

Rossant, J. and Croy, B.A. 1985. J. Embryol. exp. Morphol. 86: 177

Rubenstein, J.L.R., Nicolas, J-F. and Jacob, F. 1984. Proc. Natl. Acad. Sci. USA 81: 7137

Rubenstein, J.L.R., Nicolas, J-F. and Jacob, F. 1986. Proc. Natl. Acad. Sci. USA 83: 366

Rubin, M.R., Toth, L.E., Patel, M.D., D'Eustachio, P., Nguyen-Huu, M.C. 1986. Science 233: 663

Rubin, M.R., King, W., Toth, L.E., Sawczuk, I.S., Levine, M.S., D'Eustachio, P. and Nguyen-Huu, M.C. 1987. Mol. Cell. Biol. 7: 3836

Rudnicki, M.A., Ruben, M. and McBurney, M.W. 1988. Mol. Cell. Biol. 8: 406

Rugh, R. 1968. The Mouse; its reproduction and development. Minneapolis, Burgess.

Ruther, U., Koenen, M., Sippel, A.E. and Muller-Hill, B. 1982. Proc. Natl. Acad. Sci. USA 79: 6852

Ruther, U. and Muller-Hill, B. 1983. EMBO J 2: 1791

Ruther, U., Garber, C., Komitowski, D., Muller, R. and Wagner, E.F. 1987. Nature 325: 412

Ryskov, A.P., Ivanov, P.L., Kramerov, D.A. and Georgiev, G.P. 1983. Nuc. Acids Res. 11: 6541

Sage, H., Trueb, B. and Bornstein, P. 1983. J. Biol. Chem. 258: 13391

Sandberg, M. and Vuorio, E. 1987. J. Cell. Biol. 104: 1077

Sanger, F., Nicklen, S. and Coulson, A.R. 1977. Proc. Natl. Acad. Sci. USA 74: 5463

Sanger, F. and Coulson, A.R. 1978. FEBS Lett. 87: 107

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. 1980. J. Mol. Biol. 143: 161

Schindler, J., Matthaei, K.I. and Sherman, M.I. 1981. Proc. Natl. Acad. Sci. USA 78: 1077

Schnieke, A., Harbers, K. and Jaenisch, R. 1983. Nature 304: 315

Schnieke, A., Dziadek, M., Bateman, J., Mascara, T., Harbers, K., Gelinas, R. and Jaenisch, R. 1987. Proc. Natl. Acad. Sci. USA 84: 764

Scott, M.R.D. 1982. Ph.D. Thesis, University of London.

Scott, R.W., Vogt, T.F., Croke, M.E. and Tilghman, S.M. 1984. Nature 310: 562

Searle, A.G., Peters, J., Lyon, M.F., Evans, E.P., Edwards, J.H. and Buckle, V.J. 1987. Genomics 1: 3

Segal, S. and Khoury, G. 1979. Proc. Natl. Acad. Sci. USA 76: 5611

Sejersen, T., Sumegi, J. and Ringertz, N.R. 1985. Exp. Cell Res. 160: 19

Sejersen, T., Bjorklund, H., Sumegi, J. and Ringertz, N.R. 1986. J. Cell. Phys. 127: 274

Sekikawa, K. and Levine, A.J. 1981. Proc. Natl. Acad. Sci. USA 78: 1100

Seldin, D.C., Austen, K.F. and Stevens, R.L. 1985. J. Biol. Chem. 260: 11131

Seliger, B., Kollek, R., Stocking, C., Franz, T. and Ostertag, W. 1985. Mol. Cell. Biol. 6: 286

Shaw, G. and Kamen, R. 1986. Cell 46: 659

Sherman, M.I. 1975. in Teratomas and Differentiation. eds. Sherman, M.I. and Solter, D. A.P. New York.

Shuman, H.A., Silhavy, T.J. and Beckwith, J.R. 1980. J. Biol. Chem. 255: 168

Silbermann, M. and Frommer, J. 1972. Anat. Rec. 172: 659

Silver, J. 1985. J. Hered. 76: 436

Singer, R.H. and Ward, D.C. 1982. Proc. Natl. Acad. Sci. USA 79: 7331

Singh, K., Saragosti, S. and Botchan, M. 1985. Mol. Cell. Biol. 5: 2590

Sleigh, M.J. and Lockett, T.J. 1985. EMBO J. 4: 3831

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A. and Kucherlapati, R.S. 1985. Nature 317: 230

Sneider, T.W. and Potter, V.R. 1969. J. Mol. Biol. 42: 271

Snow, M.H.L. 1977. J. Embryol. Exp. Morphol. 42: 293

Snow, M.H.L. 1981. J. Embryol. exp. Morph. 65: (supplement) 269

Snow, M.H.L. 1985. in Molecular determinants of animal form. Liss.

Solter, D. and Knowles, B.B. 1978. Proc. Natl. Acad. Sci. USA 75: 5565

Solter, D. and Knowles, B.B. 1979. Curr. Top. Dev. Biol. 13: 139

Sonigo, P., Wain-Hobson, S., Bougueleret, L., Tiollais, P., Jacob, F. and Brulet, P. 1987. Proc. Natl. Acad. Sci. 84: 3768

Soriano, P., Gridley, T. and Jaenisch, R. 1987. Genes Dev. 1: 366

Southern, E.M. 1975. J. Mol. Biol. 98: 503

Stacey, A., Bateman, J., Choi, T., Mascara, T., Cole, W. and Jaenisch, R. 1988. Nature 332: 131

Stacey, A.J. and Evans, M.H. 1984. EMBO J. 3: 2279

Stanley, K.K. 1983. Nuc. Acids Res. 11: 4407

Steinmetz, M., Stephan, D. and Lindahl, K.F. 1986. Cell 44: 895

Stern, C.D. and Keynes, R.J. 1986. in Somites in developing embryos. eds. Bellairs, R., Ede, D.A. and Lash, J.W. Plenum Press.

Stern, C.D., Sisodiya, S.M. and Keynes, R.J. 1986. J. Embryol. exp. Morphol. 91: 209

Sternberg, M.J.E. and Thornton, J.M. 1976. J. Mol. Biol. 105: 367

Stevens, L.C. 1967. J. Natl. Cancer Inst. 38: 549

Stevens, L.C. 1970. Dev. Biol. 21: 364

Stewart, C.L., Stuhlman, H., Jahner, D. and Jaenisch, R. 1982. Proc. Natl. Acad. Sci. USA 79: 4098

Stewart, C.L., Vanek, M. and Wagner, E.F. 1985. EMBO J. 4: 3701

Stewart, C.L., Scheutze, S., Vanek, M. and Wagner, E.F. 1987. EMBO J. 6: 383

Stinnakre, M-G., Evans, M.J., Willison, K.R. and Stern, P.L. 1981. J. Embryol. Exp. Morphol. 61: 117

Stocking, C., Kollek, R., Bergholz, U. and Ostertag, W. 1985. Proc. Natl. Acad. Sci. USA 82: 5746

Stocking, C., Kollek, R., Bergholz, U. and Ostertag, W. 1986. Virology 153: 145

Stockwell, R.A. 1979. Biology of cartilage cells. C.U.P.

Stoflet, E.S., Koeberl, D.D., Sarkar, G. and Sommer, S.S. 1988. Science 239: 491

Stout, J.T. and Caskey, C.T. 1987. Methods in Enzymology. 151, in press.

Strauss, F. and Varshavsky, A. 1984. Cell 37: 889

Strickland, S. and Mahdavi, V. 1978. Cell 15: 393

Strickland, S., Smith, K.K. and Marotti, K.R. 1980. Cell 21: 347

Sudo, H., Kodama, H-A., Amagai, Y., Yamamoto, S. and Kasai, S. 1983. J. Cell Biol. 96: 191

Swank, R.T. and Bailey, D.W. 1973. Science 181: 1249

Swartzendruber, D.E. and Lehman, J.M. 1975. J. Cell. Phys. 85: 179

Tabor, S. and Richardson, C.C. 1987. Proc. Natl. Acad. Sci. USA 84: 4767

Tainer, J.A., Getzoff, E.D., Alexander, H., Houghten, R.A., Olson, A.J. and Lerner, R.A. 1984. Nature 312: 127

Takase, S., Ong, D.E. and Chytil, F. 1986. Arch. Biochem. Biophys. 247: 328

Tam, P.P.L. 1981. J. Embryol. Exp. Morphol.(supplement) 65: 103

Tam, P.P.L. and Meier, S. 1982. Am. J. Anat. 164: 209

Tan, S.S. and Morriss-Kay, G. 1985. Cell Tissue Res. 240: 403

Tan, S.S. and Morriss-Kay, G. 1986. J.Embryol. exp. Morphol. 98: 21

Taub, R.A., Hollis, G.F., Hieter, P.A., Korsmeyer, S., Waldmann, T.A. and Leder, P. 1983. Nature 304: 172

Taylor, B.A. 1978. in Origins of inbred mice. ed. Morse, H.C. pages 423-438. A.P.

Taylor, B.A. 1981. in Genetic variants and strains of the laboratory mouse. ed. Green, M.C. Fischer. Stuttgart; New York.

Taylor, B.A., Bailey, D.W., Cherry, M., Riblet, R. and Weigert, M. 1975. Nature 256: 644

Taylor, W.R. 1986. in Nucleic acid and protein sequence analysis. eds. Bishop, M.J. and Rawlings, C.J. IRL Press.

Taylor, W.R. and Thornton, J.M. 1983. Nature 301: 540

Taylor, W.R. and Thornton, J.M. 1984. J. Mol. Biol. 173: 487

Teich, N.M., Weiss, R.A., Martin, G.R. and Lowy, D.R. 1977. Cell 12: 973

Thaller, C. and Eichele, G. 1987. Nature 327: 625

Theiler, K. 1972. The house mouse. Springer, Berlin.

Thomas, K.R. and Capecchi, M.R. 1987. Cell 51: 503

Tilghman, S.M. and Belayew, A. 1982. Proc. Natl. Acad. Sci. USA 79: 5254

Timpl, R., Paulsson, M., Dziadek, M. and Fujiwara, S. 1987. Methods in Enzymology 145: 363

Toth, L.E., Slawin, K.L., Pintar, J.E. and Nguyen-Huu, M. 1987. Proc. Natl. Acad. Sci. USA 84: 6790

Treisman, R. 1985. Cell 42: 889

Tulchin, N. and Taylor, B.A. 1981. Genetics 99: 109

Upadhyay, S., Luciani, L.M. and Zamboni, L. 1979. Annls. Biol. Anim. Biochem. Biophys. 19: 1179

Utset, M.F., Awgulewitsch, A., Ruddle, F.H. and McGinnis, W. 1987. Science 235: 1379

Van der Putten, H., Botteri, F.M., Miller, A.D., Rosenfeld, M.G., Fan, H., Evans, R.M. and Verma, I.M. 1985. Proc. Natl. Acad. Sci. USA 82: 6148

Vasseur, M., Katinka, M., Herbommel, P., Yaniv, M. and Blagny, D. 1982. J. Virol. 43: 800

Velich, A. and Ziff, E. 1985. Cell 40: 705

Vilette, D., Emanoil-Ravier, R., Tobaly, J. and Peries, J. 1985. Biochem. Biophys. Res. Comm. 128: 513

Vogelstein, B. and Gillespie, D. 1979. Proc. Natl. Acad. Sci. 76: 615

Vogt, T.F., Soltor, D. and Tilghman, S.M. 1987. Science 236: 301

Vogt, T.F., Compton, R.S., Scott, R.W. and Tilghman, S.M. 1988. Nuc. Acids. Res. 16: 487

Wagner, E.F., Covarrubias, L., Stewart, T.A. and Mintz, B. 1983. Cell 35: 647

Wagner, E.F., Vanek, M. and Vennstrom, B. 1985. EMBO J. 4: 663

Wagner, E.F. and Stewart, C.L. 1986. in Experimental approaches to mammalian embryonic development. eds. Rossant, J. and Pedersen, R.A. CUP.

Wahl, G.M., Stern, M. and Stark, G.R. 1979. Proc. Natl. Acad. Sci. USA 76: 3683

Walter, G. 1986. J. Immunol. Meth. 88: 149

Walter, G., Scheidtmann, K-H., Cabonne, A., Laudano, A.P. and Doolittle, R.F. 1980. Proc. Natl. Acad. Sci. USA 77: 5197

Wang, S-Y. and Gudas, L.J. 1983. Proc. Natl. Acad. Sci. USA 80: 5880

Wang, S-Y. and Gudas, L.J. 1984. J. Biol. Chem. 259: 5899

Wang, S-Y., LaRosa, G.J. and Gudas, L.J. 1985. Dev. Biol. 107: 75

Wassarman, P.M., Greve, J.M., Perona, R.M., Roller, R.J. and Salzmann, G.S. 1984. in Molecular biology of development. eds. Davidson, E. and Firtel, R. Liss, New York.

Watson, C.J. 1986. Ph.D. Thesis, University of London.

Watson, C.J. and Jackson, J.F. 1985. in DNA cloning vol. I. ed. Glover, D.M. IRL Press.

Weber, P.C. and Salemme, F.R. 1980. Nature 287: 82

Weinstock, G.M., Rhys, C., Berman, M.L., Hampar, B., Jackson, D., Silhavy, T.J., Weisemann, J. and Zweig, M. 1983. Proc. Natl. Acad. Sci. USA 80: 4432

Wensink, P., Finnegan, D., Donelson, J.E. and Hogness, D.S. 1974. Cell 3: 315

Westaway, D., Goodman, P.A., Mirenda, C.A., McKinley, M.P., Carlson, G.A. and Prusiner, S.B. 1987. Cell 51: 651

Westhof, E., Altschuh, D., Moras, D., Bloomer, A.C., Mondragon, A., Klug, A. and Van Regenmortel, M.H.V. 1984. Nature 311: 123

White, R., Woodward, S., Leppert, M., O'Connell, P., Hoff, M., Herbst, J., Lalouel, J-M., Dean, M. and Vande Woude, G. 1985. Nature 318: 382

Wilbur, W.J. and Lipman, D.J. 1983. Proc. Natl. Acad. Sci. USA 80: 726

Wilkinson, D.G., Bailes, J.A., Champion, J.E. and McMahon, A.P. 1987a. Development 99: 493

Wilkinson, D.G., Bailes, J.A. and McMahon, A.P. 1987b. Cell 50: 79

Willison, K. and Ashworth, A. 1987. Trends Genet. 3: 351

Wingender, E., 1988. Nuc. Acids Res. 16: 1879

Wolgemuth, D.J., Gizang-Ginsberg, E., Englemyer, E., Gavin, B.J. and Ponzetto, C. 1984. Gamete Res. 12: 1

Woychik, R.P., Stewart, T.A., Davis, L.G., D'Eustachio, P. and Leder, P. 1985. Nature 318: 36

Wyman, A.R., Wolfe, L.B. and Botstein, D. 1985. Proc. Natl. Acad. Sci. USA 82: 2880

Wyman, A.R., Wertman, K.F., Barker, D., Helms, C. and Petri, W.H. 1986. Gene 49: 263

Yanisch-Perron, C., Vieira, J. and Messing, J. 1985. Gene 33: 103

Young, P.R. and Tilghman, S.M. 1984. Mol. Cell. Biol. 4: 898

Young, R.A. and Davis, R.W. 1983. Proc. Natl. Acad. Sci. USA 80: 1194

Zabarovsky, E.R. and Allikmets, R.L. 1986. Gene 42: 119