Homeobox Gene Expression in Murine Embryonic Stem Cells

A thesis submitted to the University of Adelaide

for the degree of Doctor of Philosophy

by

Paul Quinton Thomas, B.Sc. (Hons)

Department of Biochemistry

University of Adelaide

Adelaide, South Australia

October 1994

THESIS SUMMARY

Mutational and genetic analyses of homeobox gene expression and function in the post-gastrulation murine embryo have indicated that this class of transcription factors plays a key role in developmental processes such as pattern formation and specification of regional and cellular identity. Embryonic stem (ES) cells are derived from the pluripotent inner cell mass cells of the early mouse embryo and provide a model *in vitro* system for investigation of the molecular mechanisms that regulate pluripotent cell proliferation and differentiation. The overall aim of this thesis was to identify homeobox genes which may have a developmental role during early embryogenesis by the characterization of homeobox gene expression in undifferentiated ES cells, and in a range of differentiated ES cell derivatives.

PCR screening of undifferentiated ES cell cDNA identified thirteen homeobox genes, including three novel genes termed *Hesx1*, *Hesx2* and *Hesx4*. Comparison of the partial homeodomain sequences of *Hesx1*, *2* and *4* indicated that these genes did not belong to existing homeodomain sequence classes, and may therefore define novel sequence classes. Expression of these homeobox genes in ES cells and in a range of differentiated ES cell derivatives was investigated by northern blot analysis. *Hesx1*, *Hesx4*, *Hex* and *MmoxA/B* exhibited an unusual expression pattern in that they were expressed by undifferentiated ES cells and then downregulated during ES cell differentiation. Comparison between the expression of these genes and the pluripotent cell marker *Oct-4* in uninduced and induced ES cells indicated that:

i) *Hesx1* was expressed by all of the residual pluripotent cells which remain after ES cell induction;

ii) *MmoxA/B* expression was downregulated in a sub-population of pluripotent cells, indicating the existence of distinct populations of pluripotent cells that were *MmoxA/B*$^-$/*Oct-4*$^+$ and *MmoxA/B*$^+$/*Oct-4*$^+$;

iii) *Hesx4* and *Hex* expression may also define sub-populations of pluripotent cells that arise during MBA and DMSO induction. These genes were also expressed by differentiated cells induced by spontaneous and RA induction.

The expression pattern of *Hesx1*, *Hesx4*, *Hex* and *MmoxA/B* therefore points to the existence of multiple pluripotent cell types that arise during ES cell differentiation *in vitro*, perhaps as differentiation intermediates. These genes may provide useful markers for the identification of these cells *in vivo*.

Hesx1 was selected for further analysis, and cDNA clones derived from the 1.2 kb and 1.0 kb *Hesx1* transcripts were isolated. An identical 185 amino acid open reading frame was identified in both cDNA clones. The *Hesx1* open reading frame shared considerable sequence and structural homology with the *Xenopus* homeobox gene XANF-1, suggesting that these genes may be derived from a common ancestral gene. RNAase protection analysis detected *Hesx1* expression in a range of ES cell lines, which supported the expression, and perhaps function, of *Hesx1* in pluripotent cells during early embryogenesis. *Hesx1* expression was also detected in embryonic but not adult liver and the extra-embryonic amnion, which suggested that *Hesx1* may play a second role in haematopoiesis.

Isolation of a 16 kb *Hesx1* genomic clone allowed characterization of the *Hesx1* locus. Comparison between the genomic and cDNA clone sequences indicated that the alternative *Hesx1* transcripts were generated by differential usage of a poor consensus splice site within exon IV. An interesting feature of *Hesx1* was the existence of two introns within the homeobox. This arrangement has not been described for other vertebrate homeobox genes.

Physical mapping indicated that the *Hesx1* locus was on chromosome 14 bands A3-B, spanning a 7.8 cM region which contained the *Waved coat* (*Wc*) early developmental mutation. The possibility that the *Wc* mutant strain resulted from mutation in *Hesx1* was investigated by southern blot analysis of the *Hesx1* locus and sequence comparison of the *Hesx1* open reading frame in *Wc*/+ and +/+ genomic DNA. This analysis suggested that there was not an allelic relationship between *Hesx1* and the *Wc* developmental mutation.

Generation of stable ES cell lines containing the *Hesx1* open reading frame driven by an interferon inducible promoter, allowed preliminary analysis of *Hesx1* function in ES cells. No significant difference in the level of ES cell differentiation was detected in parallel cultures of uninduced and induced *Hesx1* stable ES cell monolayers cultured in a range of LIF concentrations. Also, formation and differentiation of embryoid bodies derived from

uninduced and induced stable ES cells appeared to be identical. Therefore, this analysis did not support a direct role for *Hesx1* in the maintenance of ES cell pluripotency, or in ES cell differentiation. Further analysis was therefore required to determine the developmental role of *Hesx1* in pluripotent cells.

Statement

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give my consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

SIGNED..,..... DATE....28/10/94...

ACKNOWLEDGEMENTS

I would like to thank Professor G. E. Rogers for the giving me the opportunity to undertake a Ph. D. in the Department of Biochemistry.

Many people have contributed to making the last three and a half years an enjoyable experience. Firstly, I would like to thank Dr. Peter Rathjen for his excellent supervision, unbounded enthusiasm, mateship, and for establishing the fine institution of lab golf day. Also, thanks Pete for critically reading this thesis, and for picking out my split infinitives. Thanks must also go to everyone in the Rathjen group for friendship and advise. In particular: Bryan Haines for comic genius, Wednesday nights and "corro-dobs"; "Bruce" Linda Whyatt for her caustic wit; Julie-Anne Lake for lots of laughs; Mikey B. for more "corro-dobs" and conversation; Roger Voyle for providing a different perspective.

I would also like to thank Joy Rathjen, Rachel Dempster, Jenny Washington and Sarah Retallack for the odd plate of cells.

Thanks also to Messers. H. G. Nelson and Roy Slaven for entertaining company on Saturday afternoons!

Thanks to Mum and Dad for their support, and always taking an interest in my studies.

Finally, I must thank Sarah for her understanding, encouragement and ability to keep things in perspective during the course of my Ph. D. Her unwavering emotional support has been invaluable, and I am now looking forward to our future together.

During the course of my work I was supported by a Australian Postgraduate Research Award.

CHAPTER 1: GENERAL INTRODUCTION

CHAPTER 1: GENERAL INTRODUCTION

1.1 Why Study Development?

Development is the coordinated process whereby the fertilized egg is transformed into the complex structures which constitute the mature organism. The ambition to understand this complex and fascinating process has driven the study of developmental biology since the early 1900's when many of the basic principles were established. Now, with the advent of molecular biology, we are approaching an understanding of the underlying molecular interactions which control development.

The study of mammalian development is perhaps the most interesting and relevant of all developmental biology. Firstly, mammalian development is intrinsically interesting because it provides insight into the generation of the human form. Secondly, on a more practical level, investigation and understanding of mammalian development using a molecular genetic approach has the potential to contribute to the establishment of animal models for human disease, and to improve productivity of agricultural animals. However, the investigation of mammalian development is difficult because of the large genome size, the relatively long generation time and small litter size. Moreover, the embryo develops inside the mother and is therefore relatively inaccessible for experimental manipulation.

1.2 Murine Development

Despite these inherent difficulties, the mouse has become established as the primary experimental animal for the study of mammalian development. There are several factors which combine to make the mouse a powerful model system. Firstly, almost 100 years of genetical research has located over 700 genes, many of which affect murine development. Secondly, through the use of electron microscopy, a detailed knowledge of the morphological changes that occur during murine embryogenesis has been established. Thirdly, the ability to isolate and culture pluripotent cells has made possible the analysis of developmental signals and genes which direct cellular differentiation and proliferation in the early mouse embryo. Furthermore, the capacity to introduce specific genetic mutations into

pluripotent stem cells *in vitro*, coupled with the ability to reintroduce these cells into the blastocyst, allows direct analysis of developmental gene function *in vivo*. Therefore, through a molecular genetic approach, it is now possible to identify and unravel the molecular mechanisms which co-ordinate mammalian ontogeny.

1.2.1 Pre-implantation Development

Murine development begins with the fertilization of the oocyte with the sperm. During fertilization the sperm penetrates the outer layers of the oocyte (cumulus mass and zona pelucida) and enters the cytoplasm, triggering the second meiotic division of the oocyte. Nuclear membranes form around the maternal and paternal chromosomes, and the haploid pronuclei move toward the centre of the egg. At this stage DNA replication takes place and, after breakdown of the nuclear membrane, the chromosomes assemble on the spindle.

Shortly after chromosome alignment, the first cleavage event of the embryo takes place. The embryonic cells continue to divide approximately every 20 hours, to give rise to the eight cell embryo (Fig. 1.1). It is during this period that the embryo undergoes the oogenic-embryonic transition, where the maternal genetic instructions (i.e. maternal mRNA secreted into the oocyte) are superseded by the products of embryonic transcription (Kidder, 1992). This occurs at the 2 cell stage where there is a dramatic decline in oogenic mRNA followed by a general transcriptional activation of the embryonic genome (Latham *et al.*, 1991).

Up to the early eight cell embryo stage, it appears that the individual embryonic cells, termed blastomeres, are totipotent. This has been shown by extirpation of one of the two cell stage blastomeres, which resulted in normal development (Tarkowski, 1959). Similarly, blastomeres from the four cell stage embryos can also give rise to the entire mouse (Hogan *et al.*, 1986).

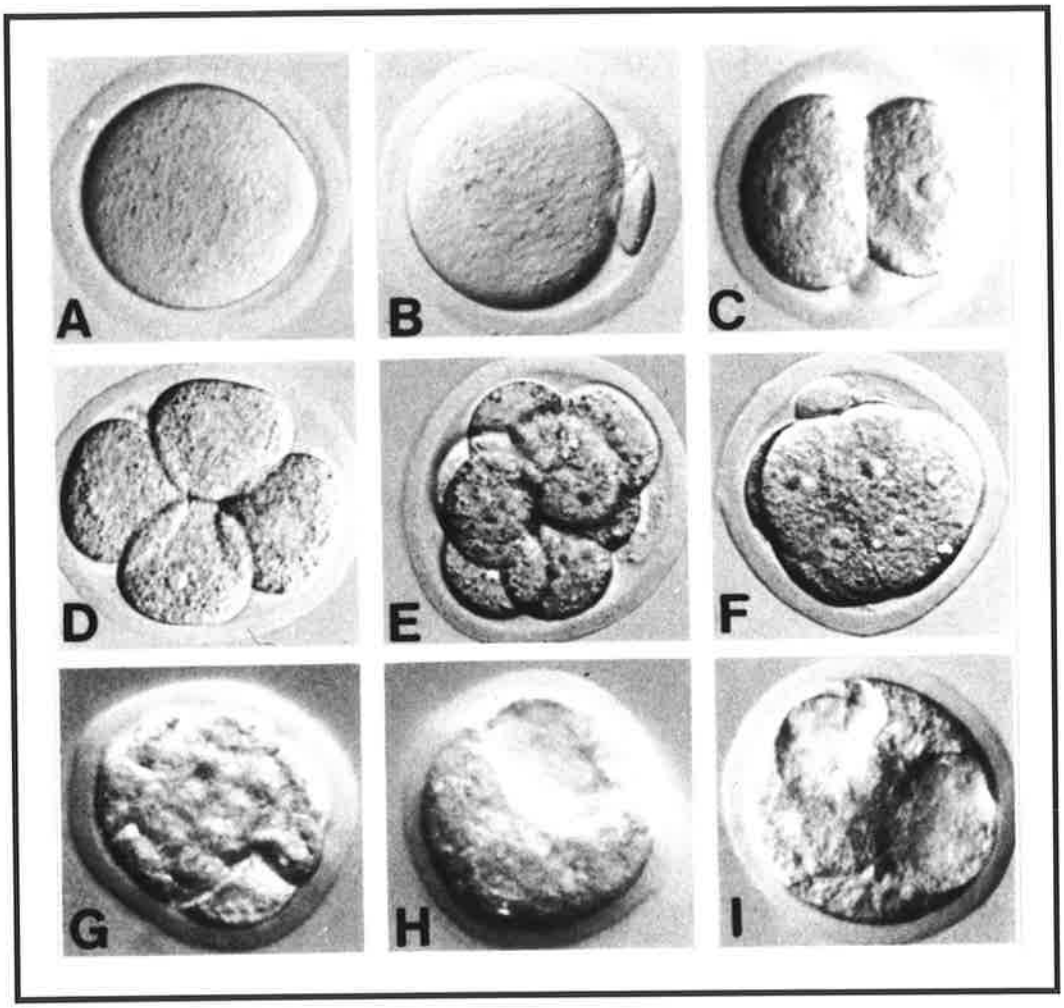
1.2.2 Blastocyst Formation

After formation of the eight cell embryo at approximately 60 hours post-fertilisation, embryonic compaction occurs (Fig. 1.1E). At this stage the individual blastomeres flatten

Figure 1.1 Development of the preimplantation murine embryo.

(A) Oocyte arrested in metaphase II of meiosis. (B) The fertilized egg. (C) The two cell stage embryo. (D) The four cell stage embryo. (E) The eight cell stage embryo before compaction. (F) The late eight cell stage embryo after compaction. (G) The 16 cell stage embryo. (H) The 32 cell stage embryo. (I) The 32 cell stage embryo showing the onset of cavitation.

Adapted from Maro *et al.* (1991).


A
B
C
D
E
F
G
H
I

upon each other to maximise cell contact, establish gap-junctional coupling, and develop a polar phenotype both cytocortically and cytoplasmically.

This establishment of cellular polarity is believed to be the basis for the first partitioning in cellular potential which occurs during embryogenesis (Johnson and Maro, 1988). As the embryo continues to divide, cells at or near the centre of the embryo form the inner cell mass, which are the pluripotent cells from which the embryo proper is derived (Fig. 1.2A). The apical blastomeres pursue the trophectoderm cell lineage (Pederson, 1988), which forms an outer layer of epithelium surrounding the ICM and the fluid filled cavity (blastocoel), and gives rise to extra-embryonic tissue including the chorion and the placenta.

At 4 days post coitum (p.c.), shortly before implantation, a second developmental partitioning occurs. The ICM cells which line the blastocoel cavity differentiate into an epithelial layer termed the primitive endoderm (Fig. 1.2B). This lineage will eventually give rise to the extra-embryonic parietal and visceral endoderm in the yolk sacs which surround the developing embryo. At 4-5 days p.c. the mature blastocyst, which contains about 64 cells of which 20 are located in the ICM, hatches from the surrounding zona pelucida and is ready to implant into the uterine wall.

1.2.3 Post-implantation Development

Development of the pre-gastrulation embryo after implantation is shown in Fig. 1.3. Implantation of the embryo is mediated by the mural trophectoderm cells which first invade the uterine epithelium and the underlying basal lamina and then penetrate the uterine stroma. At 5.5-6 days p.c., the primitive ectoderm cells (which form the embryo proper), undergo rapid division and flatten to form a layer of epithelium supported by a basal lamina laid down between the ectoderm and the visceral endoderm. The proamniotic cavity forms within the primitive ectoderm and expands as the primitive ectoderm cells flatten.

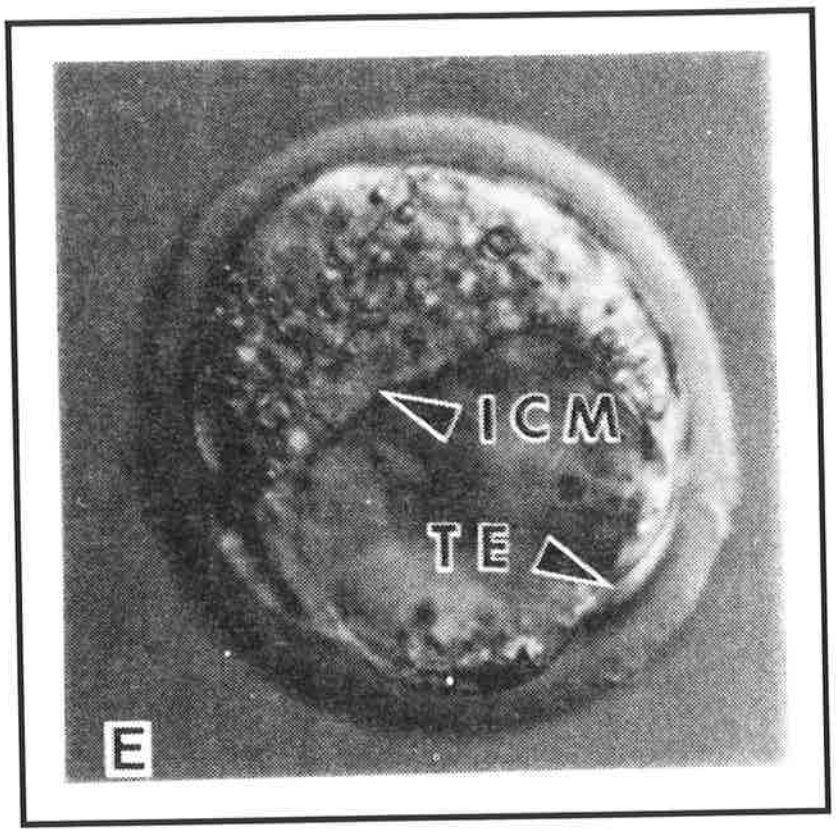
Two distinct extra-embryonic endoderm lineages derived from the primitive endoderm also appear during this period. The parietal endoderm cells, which will eventually form the parietal yolk sac, are derived from primitive endoderm cells which migrate onto the inner surface of the trophectoderm. The visceral endoderm, from which the visceral yolk sac is

Figure 1.2 The blastocyst stage embryo.

A. An early blastocyst stage embryo (approximately 32 cells) with discernible inner cell mass (ICM) and trophectoderm (TE) cells. **B.** A section of the mouse blastocyst at 4.5 days of development showing the inner cell mass (epiblast), primitive endoderm (PrEnd) and trophectoderm (TE) cells.

A. Reproduced from Pederson (1988); B. Reproduced from Hogan *et al.* (1986).

A



B

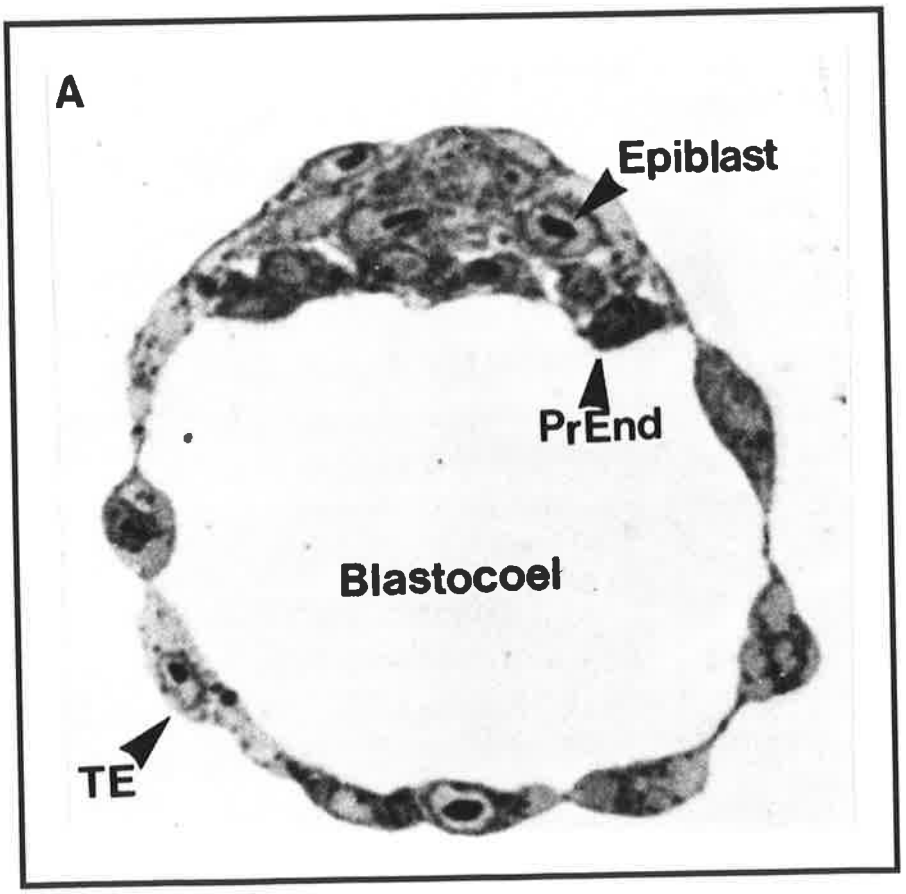
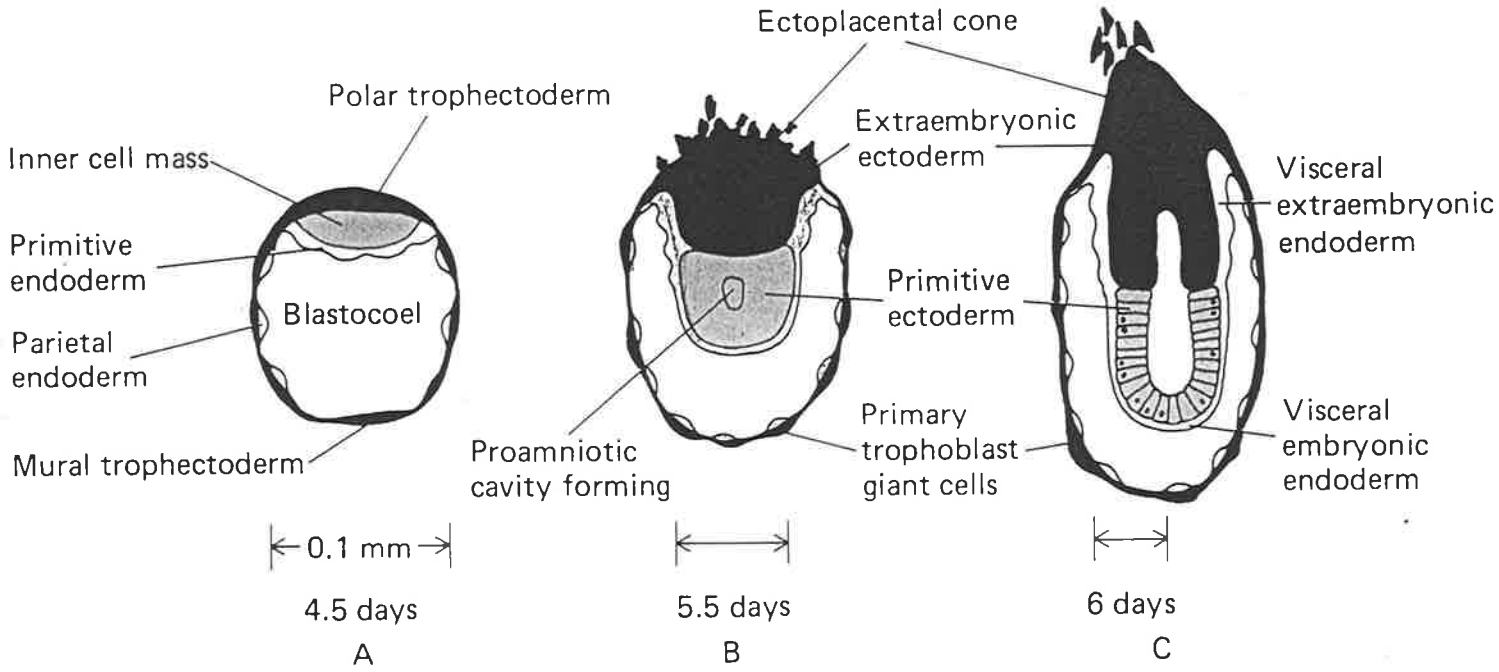


Figure 1.3 Early postimplantation murine embryogenesis.

Schematic representation of the postimplantation stages of murine development. The pluripotent tissues are shaded. **A.** At 4.5 days p.c. (post-coitum) the embryo consists of the inner cell mass cells surrounded by the primitive endoderm and trophectoderm cell lineages. **B.** By 5.5 days p.c. the proamniotic cavity begins to form within the early primitive ectoderm cells. The pluripotent primitive ectoderm cells are formed from the inner cell mass cells. **C.** At 6 days p.c. the primitive ectoderm flattens and expands to form an epithelial layer of pluripotent cells.

Adapted from Hogan *et al.* (1986).


Polar trophectoderm
Inner cell mass
Primitive endoderm
Blastocoel
Parietal endoderm
Mural trophectoderm
0.1 mm
4.5 days
A
Ectoplacental cone
Extraembryonic ectoderm
Primitive ectoderm
Proamniotic cavity forming
Primary trophoblast giant cells
5.5 days
B
Visceral extraembryonic endoderm
Visceral embryonic endoderm
6 days
C

derived, is comprised of two morphologically distinct cell populations: the embryonic visceral endoderm which lines the primitive ectoderm and the extra-embryonic visceral endoderm which contacts the extraembryonic ectoderm.

1.2.4 Gastrulation

At 6.5 days p.c. the primitive ectoderm has expanded into an epithelium comprised of approximately 1000 cells (Fig. 1.3C; Hogan *et al.*, 1986), which give rise to the entire embryo, including the germ-line (Beddington, 1988). At this stage of development, gastrulation occurs which results in the generation of the mesoderm and endoderm germ-layers, with the concomitant establishment of the basic body plan of the foetus. The appearance of the primitive streak marks the beginning of gastrulation (Fig. 1.4A), and defines the anterior-posterior axis of the embryo. During gastrulation, the primitive ectoderm cells within the region of the primitive streak lose their epithelial conformation, and invaginate to produce mesoderm (Fig. 1.4B). This mesoderm layer, which accumulates between the primitive ectoderm and the visceral endoderm, migrates axially and laterally toward the anterior pole. By 8 days p.c. mesoderm migration has reached the anterior pole of the embryo resulting in a layer of mesoderm spanning the length of the embryo (Fig. 1.4C). This mesoderm tissue condenses into 65 paired blocks, or somites, along the anterior-posterior axis, which will give rise to the vertebrae, ribs, muscles and dermis of the skin.

The embryonic or definitive endoderm, from which the gut, liver and lung tissue is derived, is also induced during gastrulation. Cell labelling studies have indicated that primitive ectoderm cells migrate through the anterior part of the primitive streak and displace the primitive endoderm to form the precursor of the gut (Tam and Beddington, 1992). A smaller proportion of the definitive endoderm cells are also believed to delaminate directly from the primitive ectoderm without prior migration through the primitive streak (Tam and Beddington, 1992).

A summary of the cell lineages which give rise to the embryonic and extra-embryonic tissues is shown in Figure 1.5.

Figure 1.4 Later postimplantation murine embryogenesis.

Schematic representation of murine development from 7 days p.c. (post-coitum) to 8.5 days p.c.. **A.** At 6.5-7 days p.c. the onset of gastrulation is marked by the formation of the primitive streak at the posterior end of the primitive ectoderm. **B.** The nascent mesoderm moves anteriorly and laterally to generate a distinct layer of embryonic mesoderm. **C.** At 8.5 days p.c., a range of differentiated cell types have formed within the somites, notochord, neural fold, gut and heart precursor tissues.

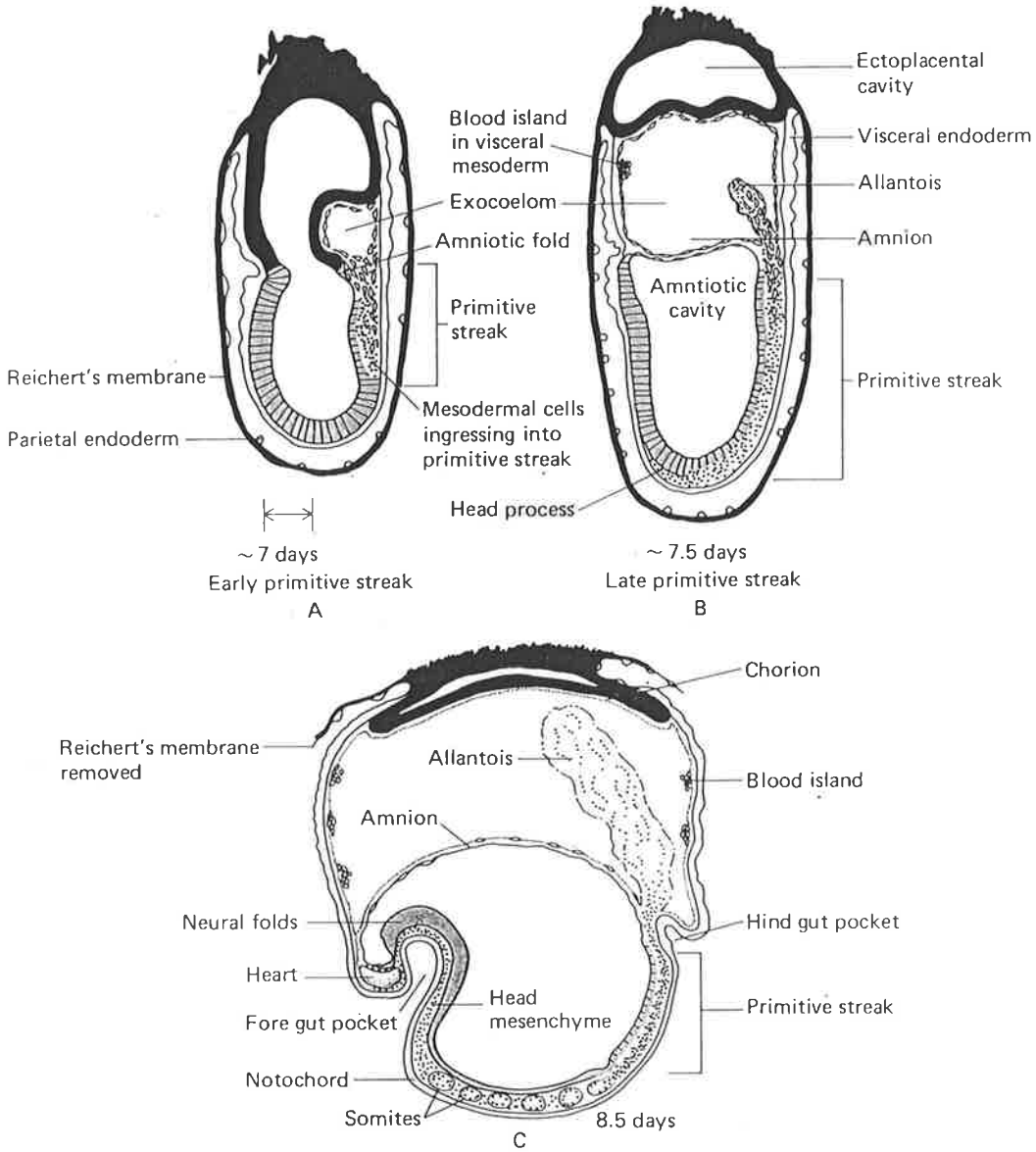
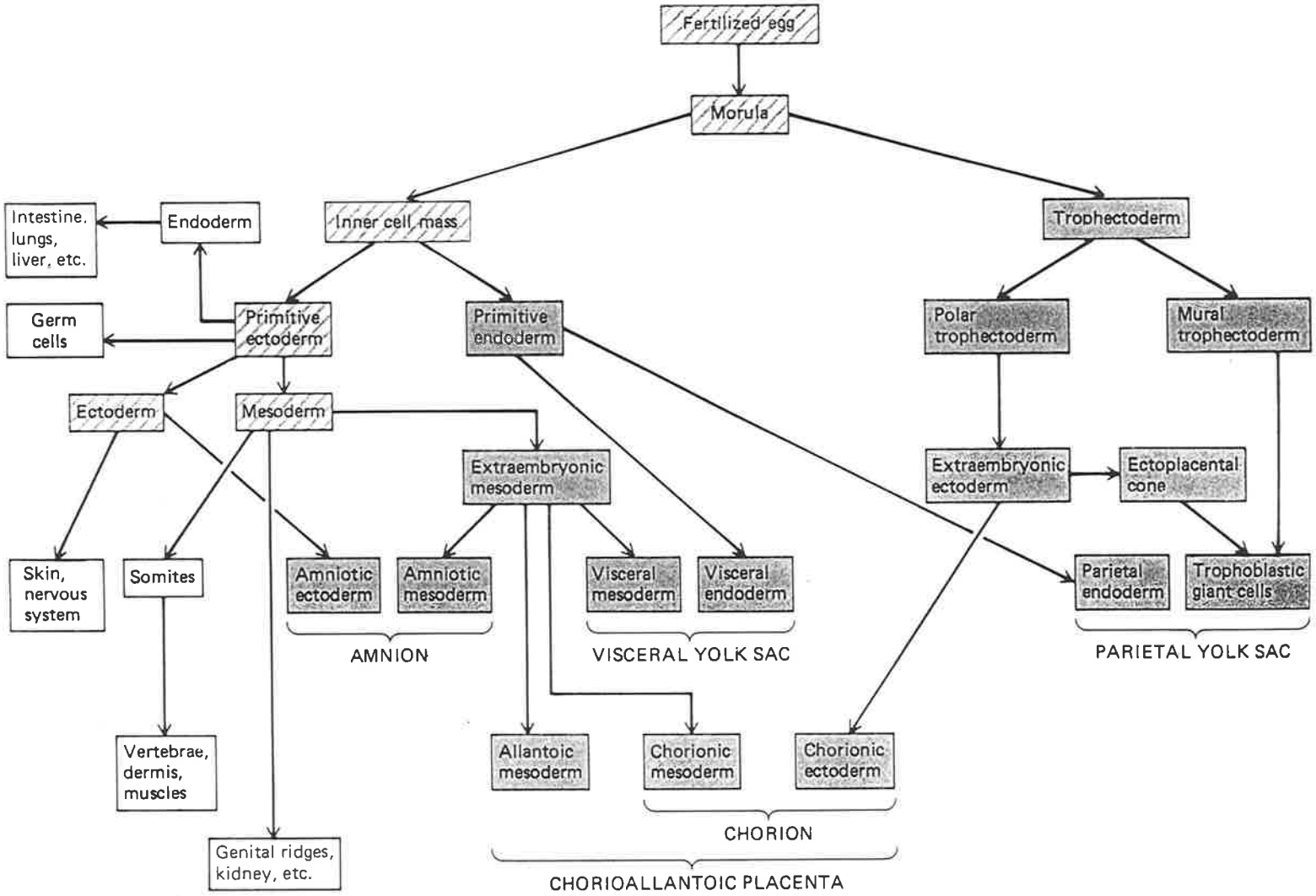

Ectoplacental cavity
Blood island in visceral mesoderm
Visceral endoderm
Allantois
Exocoelom
Amnion
Amniotic fold
Amntiotic cavity
Primitive streak
Primitive streak
Reichert's membrane
Mesodermal cells ingressing into primitive streak
Parietal endoderm
Head process
~ 7 days
Early primitive streak
A
~ 7.5 days
Late primitive streak
B
Chorion
Reichert's membrane removed
Allantois
Blood island
Amnion
Neural folds
Hind gut pocket
Heart
Head mesenchyme
Primitive streak
Fore gut pocket
Notochord
Somites
8.5 days
C

Figure 1.5 Summary of the lineages of tissues which constitute the murine embryo.

The open boxes indicate tissues of the embryo proper. The striped boxes represent tissues that will give rise to the embryo proper and the extraembryonic cells. The shaded boxes indicate extraembryonic tissues.


Fertilized egg
Morula
Inner cell mass
Trophectoderm
Endoderm
Intestine, lungs, liver, etc.
Germ cells
Primitive ectoderm
Primitive endoderm
Polar trophectoderm
Mural trophectoderm
Ectoderm
Mesoderm
Extraembryonic mesoderm
Extraembryonic ectoderm
Ectoplacental cone
Skin, nervous system
Somites
Amniotic ectoderm
Amniotic mesoderm
Visceral mesoderm
Visceral endoderm
Parietal endoderm
Trophoblastic giant cells
AMNION
VISCERAL YOLK SAC
PARIETAL YOLK SAC
Vertebrae, dermis, muscles
Allantoic mesoderm
Chorionic mesoderm
Chorionic ectoderm
Genital ridges, kidney, etc.
CHORION
CHORIOALLANTOIC PLACENTA

1.2.5 Generation of the Primordial Germ Cells during Development

A pluripotent cell lineage is maintained throughout development in the form of the primordial germ cells. These cells, which differentiate from the cells of the primitive ectoderm, are first identified at 8 days p.c. within a region towards the posterior end of the primitive streak (Snow 1981, Snow and Monk, 1983). As embryogenesis proceeds, the primordial germ cells embark upon a complex migratory pattern, which finally results with the colonization of the genital ridge of the presumptive gonad around 13.5 days p.c.. During this migration, the germ cells divide once every 16 hours. It has been estimated that the primordial germ cell population increases from 10-100 cells to approximately 25,000 cells by the time the genital ridges are fully colonized (Tam and Snow, 1981). At 12.5 days p.c., female germ cells enter meiosis in response to a stimulus provided by the somatic cells of the genital ridge (Schultz, 1988). In contrast, male germ cells enter mitotic arrest, and do not enter meiosis until later in development (Monk and McLaren, 1981).

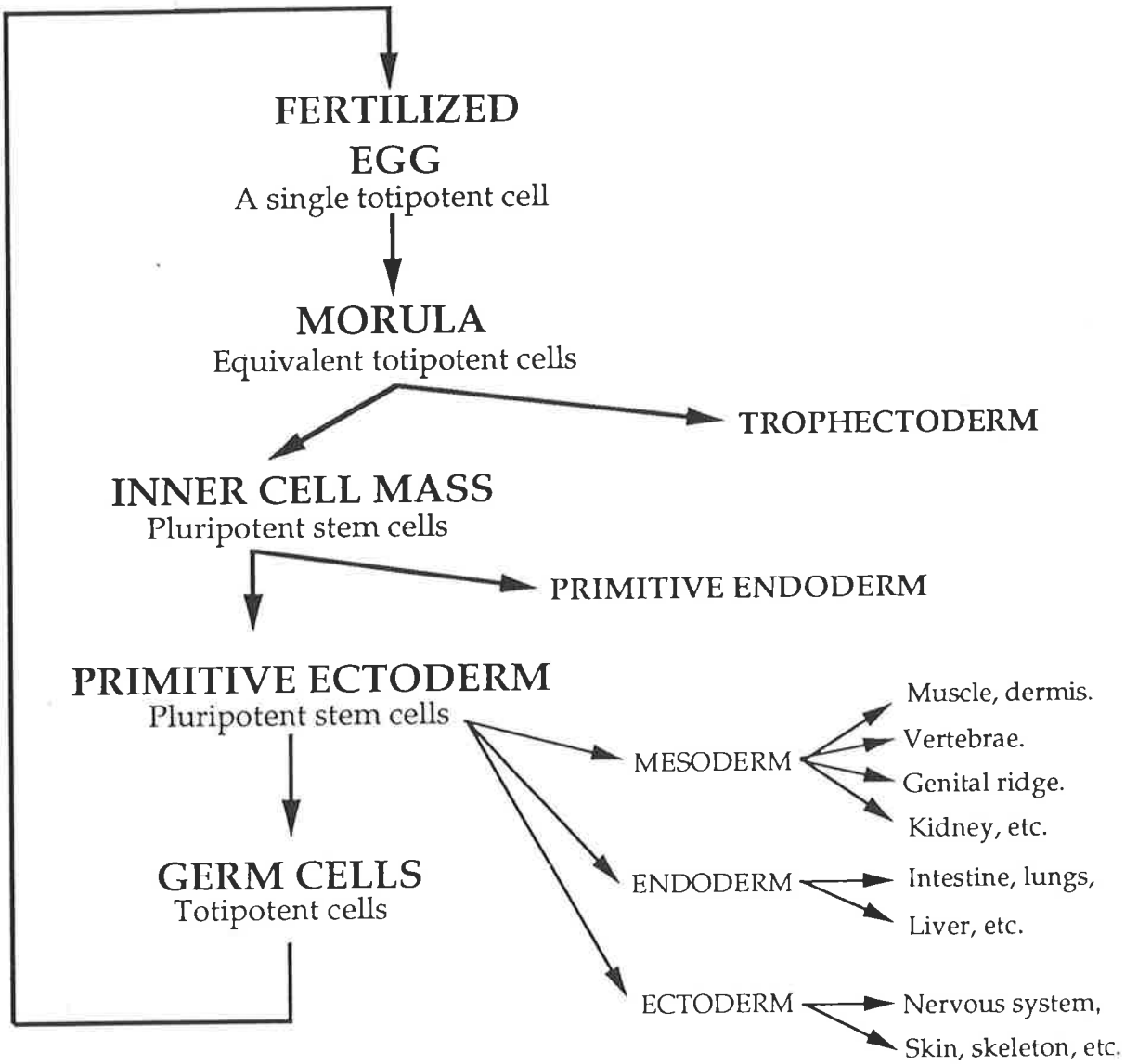
1.3 Pluripotent Cells in Early Mammalian Development

In mammalian development there is no evidence to suggest that positional information in the early embryo is laid down by maternally derived determinants, as occurs in some lower eukaryotes and vertebrates. Instead, through division of a single totipotent founder cell, the fertilised egg, a population of pluripotent cells is maintained throughout development (Fig. 1.6). The pluripotent cell lineage is established with the formation of the inner cell mass (derived from the blastomeres), is expanded in the primitive ectoderm, and is continued through the germline. *In vitro* studies have suggested that these pluripotent cell types may be interconverted via a common pluripotent intermediate (Matsui *et al.*, 1992; Rossant, 1993).

Formation of a normal mammalian embryo requires tight control of the signaling pathways which regulate proliferation and differentiation of the pluripotent stem cell pool. However, in this regard, the mammalian embryo has remarkable plasticity. This is exemplified by the ability to generate a single viable foetus of normal size from an aggregation of several pre-implantation mouse embryos . Conversely, killing up to 80% of the cells in the blastocyst or egg cylinder can still result in the formation of a normal embryo

Figure 1.6 Totipotent and pluripotent cells in murine development.

Murine development begins with a single totipotent cell, the fertilized egg. Cleavage of the fertilized egg generates the morula, which consists of equivalent totipotent cells termed blastomeres. The pluripotent cell lineage is established by the inner cell mass cells and is expanded during the formation of the primitive ectoderm. The germ cells maintain the pluripotent cell lineage for the next generation. The differentiated tissues which arise from each stage of the totipotent/pluripotent lineage are shown.


FERTILIZED EGG
A single totipotent cell
MORULA
Equivalent totipotent cells
TROPHECTODERM
INNER CELL MASS
Pluripotent stem cells
PRIMITIVE ENDODERM
PRIMITIVE ECTODERM
Pluripotent stem cells
GERM CELLS
Totipotent cells
MESODERM
Muscle, dermis.
Vertebrae.
Genital ridge.
Kidney, etc.
ENDODERM
Intestine, lungs,
Liver, etc.
ECTODERM
Nervous system,
Skin, skeleton, etc.

(Snow and Tam, 1979). This adaptive capacity of the murine embryo appears to be a function of the ability of the pluripotent stem cell pool to reprogram their developmental fate in response to a flexible signalling system. This also suggests that an understanding of murine development can be achieved by investigation of pluripotent cell response, i.e. differentiation versus proliferation, to the developmental signals within the embryo.

1.4 Embryonic Stem (ES) Cells- an *In Vitro* Model for Murine Development

Through the use of molecular biology, it is now possible to approach an understanding of the molecular interactions which constitute the developmental programme. However, the direct biochemical analysis of the embryo itself is extremely difficult due to its complex cellular nature and small size. Furthermore, the embryo develops within the uterus and is therefore relatively inaccessible to experimental manipulation. An *in vitro* model system is therefore required for molecular investigation of pluripotent cell differentiation and proliferation during murine development.

1.4.1 The Limitation of Embryonal Carcinoma (EC) Cells

Much of the early functional analysis of pluripotent cells was carried out using embryonal carcinoma (EC) cells. EC cells are derived from the stem cells of complex tumours termed teratocarcinomas, which contain a variety of endodermal, ectodermal and mesodermal differentiated tissues. EC cells are similar to the inner cell mass and primitive ectoderm cells in cell morphology, cell surface antigen expression, ultrastructure and protein synthesis profiles (Martin, 1975; Martin, 1980). EC cells are multipotent and can give rise to a variety of differentiated cell types *in vitro* and *in vivo*. However, apart from a single report (Stewart and Mintz, 1981), EC cells appear unable to colonize the germ line upon introduction to the blastocyst. This indicates that EC cells are not a "normal" pluripotent cell, and therefore have limited use as a *in vitro* model system for the investigation of pluripotent cells.

1.4.2 Isolation and Differentiation Capacity of ES Cells

Embryonic stem (ES) cell lines were first established 13 years ago from *in vitro* culture of mouse blastocysts (Evans and Kaufman, 1981) or from immunosurgically isolated inner cell mass cells (Martin, 1981). ES cells are morphologically similar to EC cells, and can be maintained in culture for extended periods in the undifferentiated state. ES cells can also be induced to differentiate along alternative pathways by exposure to chemical inducers or withdrawal of the cytokine LIF (1.4.4). An important property of ES cells is that, upon reintroduction into the blastocyst, ES cells have the ability to contribute to of all of the tissues which constitute the developing embryo, including the germline (Bradley *et al.*, 1984; Robertson, 1987). This demonstrates that ES cells cultured *in vitro* retain their full biological activity, and therefore provide a good model system for the investigation of stem cell differentiation and renewal during murine embryogenesis.

1.4.3 *In Vitro* Differentiation of ES Cells

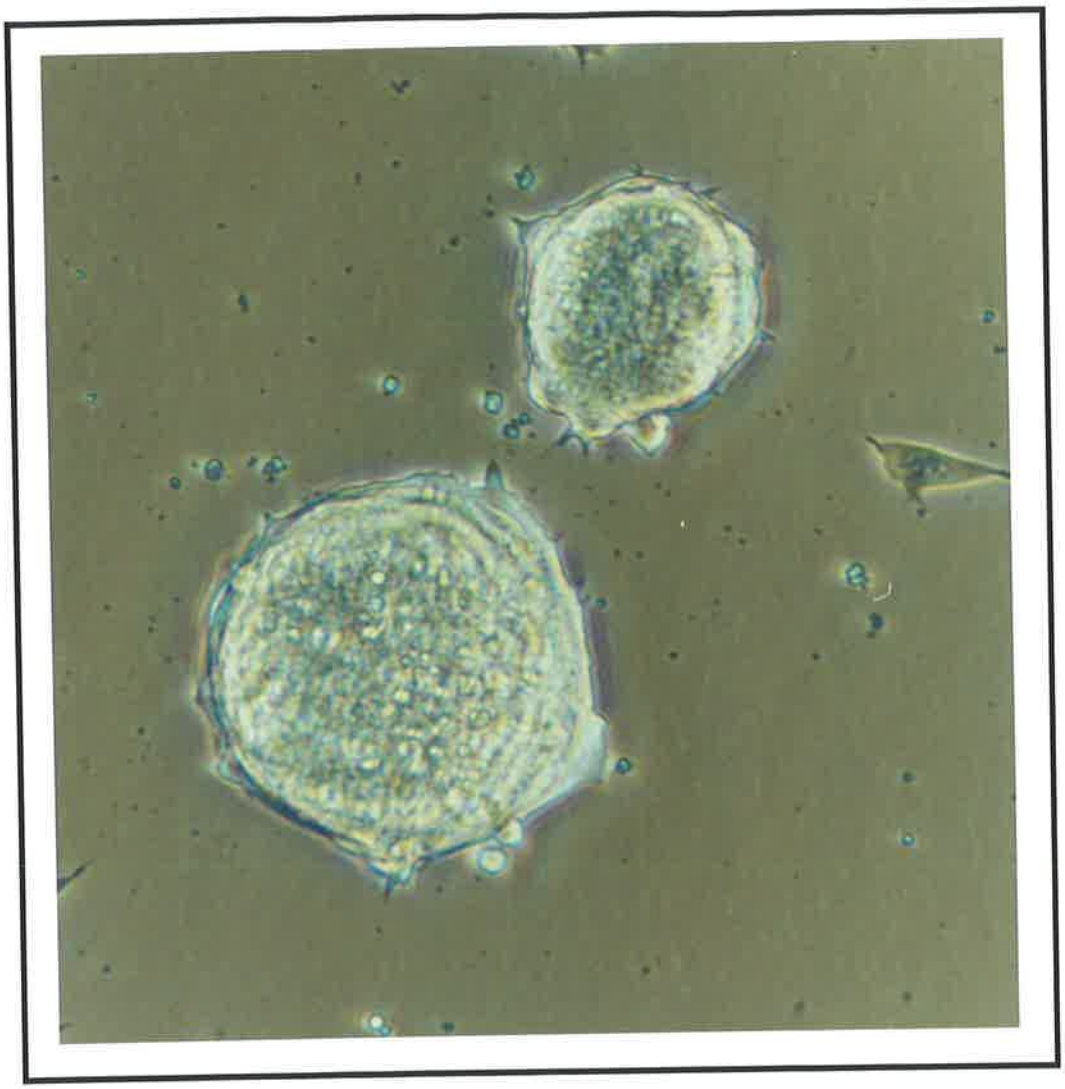
ES cells cultured in the presence of the cytokine LIF grow as dome-shaped colonies consisting of small round cells with relatively large nuclei and minimal cytoplasm (Fig. 1.7). Apart from this distinctive morphology, undifferentiated ES cells can be readily distinguished from differentiated ES cells by their expression of the pluripotent cell markers *Oct-4* (1.8), alkaline phosphatase (Adamson, 1988) and SSEA-1 (Solter and Knowles, 1978).

1.4.3.1 Chemical and Spontaneous Induction of ES Cell Monolayers

Spontaneous and chemical induction of ES cells monolayers *in vitro*, gives rise to a variety of differentiated cell types, which share characteristics with differentiated tissues within the embryo (Smith, 1992). ES cells cultured in the absence of LIF (spontaneous induction; 1.4.4) generate a range of terminally differentiated cells of ill-defined phenotype (Fig. 3.4A). These terminally differentiated cells are often found surrounding a nest of undifferentiated ES cells. The maintenance of this residual population of undifferentiated ES cells is thought to be due to feedback expression of LIF by the differentiated ES cells (Rathjen *et al.*, 1990). Retinoic acid (RA) is the most efficient of the chemical inducers.

Figure 1.7 Embryonic stem cell colony morphology.

Embryonic stem (ES) cells cultured in the presence of LIF grow as compact dome shaped colonies. ES cells were cultured for three days and were photographed under phase contrast optics at 50 X magnification.



Exposure of ES cells to RA generates two different terminally differentiated cell types; large fibroblast cells and variable amounts of refractile parietal yolk sac-like cells, which may represent extra-embryonic mesoderm (Fig. 3.4B). Under optimal conditions, treatment of ES cells with 3-methoxybenzamide (MBA) induces differentiation into a relatively uniform layer of flattened epithelial cells (Fig. 3.4C). Dimethylsulfoxide (DMSO) is a relatively inefficient differentiating agent and induces ES cell differentiation into a variety of ill defined cell types (Fig. 3.4D).

1.4.3.2 Generation of Embryoid Bodies from ES Cell Suspension Cultures

ES cells may also be cultured in suspension to form globular structures termed embryoid bodies, which share structural similarities with the early murine embryo (Robertson, 1987). ES cell aggregates cultured from 2-4 days, form simple embryoid bodies. These structures contain a central mass of undifferentiated cells (which resemble the inner cell mass) and an outer layer of endoderm (which shares similarities with the primitive endoderm). Upon further culture, cystic embryoid bodies are generated which contain an additional layer of ectoderm cells, which resembles the primitive ectoderm. Cystic embryoid bodies can be induced to differentiate into a variety of structures which resemble differentiated tissues of the post-gastrulation embryo such as beating muscle, neurons and blood islets (Doetschman *et al.*, 1985). Differentiation of ES cells in the form of embryoid bodies therefore provide an accessible *in vitro* system to study pluripotent cell differentiation in a complex "embryo-like" environment.

1.4.4 Identification of Cytokines which Inhibit ES Cell Differentiation *In Vitro*

Historically, maintenance of ES cells in the undifferentiated state required co-culture of ES cells with feeder layers of mitotically inactivated fibroblasts (Martin and Evans, 1975). The discovery of soluble factors which inhibit ES cell differentiation stemmed from the observation that differentiation was suppressed when ES cells were cultured in conditioned media from Buffalo rat liver (BRL) cells, in the absence of feeder cells (Smith and Hooper,

1987; Smith *et al.*, 1988). Purification of this differentiation inhibiting activity (DIA) showed it to be identical to a previously described cytokine, Leukaemia Inhibitory Factor (LIF; Smith *et al.*, 1988; Williams *et al.*, 1988). Since the identification and purification of LIF, other cytokines including Ciliary neurotrophic factor (CNTF) and Oncostatin M (OSM) have also shown to be able to suppress ES cell differentiation *in vitro* (Conover *et al.*, 1993; Piquet-Pellorce *et al.*, 1994). These factors appear to act through a common signal transduction pathway, via the gp130 receptor subunit, to maintain ES cells in the undifferentiated state (Taga *et al.*, 1992). The identification and purification of LIF has therefore allowed the continuous propagation of undifferentiated ES cells for extended periods as monocultures.

The ability to culture ES cells in the absence of additional feeder cells, and the ability to discriminate between undifferentiated and differentiated ES cells, allows direct investigation of ES cell differentiation and proliferation in response to soluble growth factors and chemical inducers. In addition, the complete integration of ES cells into host embryos indicates that cultured ES cells are capable of responding to the full repertoire of developmental signals which direct murine embryogenesis. Factors which regulate ES cell proliferation and differentiation *in vitro* are therefore likely to play a developmental role in pluripotent cells during early embryogenesis. The maintenance of ES cell pluripotency *in vitro*, coupled with the ability to culture ES cells as monolayers, therefore allow ES cells to be exploited as an *in vitro* model system for the identification and characterization of such regulators.

1.4.5 ES Cells and Transgenesis

Generation of transgenic mice via zygotic injection is limited by an inability to control the number of copies, and the genomic location, of transgenes (Kuehn *et al.*, 1987). The capacity of ES cells to colonize the germline has been exploited for the transmission of defined genetic information into the mouse genome. This powerful method of transgenesis, termed gene targeting, involves generation of a specific mutation at a desired locus by homologous recombination in ES cells (Capecchi, 1989; Schwartzberg *et al.*, 1989). The gene targeting approach can therefore be used for the direct analysis of developmental gene function by the generation of embryos carrying null mutant alleles. In addition, generation of

alleles carrying subtle mutations allows investigation of the relationship between protein structure and function within an *in vivo* context.

Pre-determined genetic mutations are generated by homologous recombination between the endogenous locus and a targeting vector encoding the mutation. The targeting vector also encodes a selectable marker gene, typically Neo^R, to allow for selection of cells which have incorporated the targeting vector into their genome. As homologous recombination is generally a relatively rare event compared with illegitimate recombination at random sites within the genome, a cell culture system is required to screen and isolate cells encoding the desired mutation. This is provided by ES cells which are then reintroduced into a host blastocyst to generate chimeric mice. Use of a host blastocyst strain which has a different coat colour marker from the ES cells, allows screening of chimerae for germline transmission of the desired mutation. Embryos which are homozygous for the mutated allele are then generated by intercrossing heterozygous mice derived from the germline chimeras.

1.4.6 The Relationship between ES cell and Embryonic Pluripotent Cells

The exact relationship between ES cells and pluripotent cell lineages such as the inner cell mass, primitive ectoderm and the primordial germ cells remains unclear. The isolation of ES cells from the inner cell mass, coupled with the ability of ES cells to colonize the extraembryonic endoderm and trophoblast (Beddington and Robertson, 1989), suggest that ES cells are closely related to the inner cell mass cells. However, ES cells have a very limited potential to generate trophectoderm and parietal endoderm cells in monolayer culture (Smith, 1992), and in this respect are more closely related to the primitive ectoderm. Morphologically, ES cells also resemble more closely the primitive ectoderm cells than the larger, more rounded cells of the inner cell mass. One possible explanation for this apparent contradiction is that ES cells lines do not have a direct embryonic equivalent, but rather, may represent a pluripotent cell which has been diverted from its normal developmental fate and adopted a program of continued proliferation (Robertson and Bradley, 1988). As isolation of ES cells is restricted to blastocyst embryos cultured *in vitro* for four to six days, it appears that only the embryonal pluripotent cells which are present within this time period can

undergo the proposed transition into an ES cell phenotype (with existing culture conditions). Furthermore, it is also of interest that primordial germ cells (PGC) cultured in the presence of the leukemia inhibitory factor (LIF), steel factor (SF) and basic fibroblast growth factor (bFGF) are able to form colonies which appear to be identical to ES cells in morphology and pluripotency (Matsui *et al.*, 1992). This suggests that PGCs may be able to dedifferentiate into pluripotent cell type, perhaps equivalent to a primitive ectoderm cell, which allows their subsequent transition into an ES cell.

1.4.7 ES Cells are a Source of Developmental Control Genes

ES cells injected into the blastocyst embryo have the potential to participate in the formation of all embryonic tissues, including the germline. This demonstrates that ES cells cultured *in vitro* retain the the ability to generate and respond to all the developmental signals which co-ordinate stem cell differentiation and renewal during early embryogenesis. ES cells are therefore a readily accessible resource for the isolation of developmental control genes which regulate the earliest developmental decisions during mammalian ontogeny.

1.5 Homeobox Genes

A recurring theme in the molecular investigation of transcriptional regulation has been the identification of families of transcription factors encoding highly conserved sequence motifs which mediate specific protein/DNA interaction. The structural motifs encoded by these families include the homeobox (Scott *et al.*, 1989), the zinc finger (Kaptein, 1991) and the leucine zipper (Kerppola and Curran, 1991). One of these gene families, the homeobox genes, have been shown to encode key regulatory molecules which control cellular differentiation and proliferation in many, if not all, eukaryotic species.

1.5.1 Identification of Homeobox Genes

The homeobox genes were originally identified in the fruit fly *Drosophila melanogaster* as genes responsible for inducing homeotic transformations. Homeosis is the process which results in the transformation of structures from one body segment into structures derived

from another segment (Bateson, 1894). For example, mutation at the *Antennapedia* locus induces the formation of a pair of legs in place of the antenna (Frischer *et al.*, 1986). Cloning and chromosomal localization of the *Antennapedia* gene showed that it had homology with a neighbouring pair rule gene *fushi tarazu* (*ftz*) and the homeotic *Ultrabithorax* (*Ubx*) gene (McGinnis *et al.*, 1984a; McGinnis *et al.*, 1984b). Sequencing of the region of cross-homology revealed a highly conserved 180 bp DNA segment named the homeobox (McGinnis *et al.*, 1984a; McGinnis *et al.*, 1984b; Scott and Weiner, 1984). Using the homeobox sequence as a probe, many homeotic genes and other genes involved in the control of *Drosophila* development were rapidly isolated.

It became quickly apparent that the homeobox sequences were not restricted to *Drosophila* developmental genes, but are found in all eukaryotic species including mice (McGinnis *et al.*, 1984c; Scott *et al.*, 1989; Kappen *et al.*, 1993). In fact, homeobox genes constitute one of the largest known genes families, and at least 337 homeobox genes have been isolated (Kappen *et al.*, 1993). The remarkable evolutionary conservation of the homeobox motif suggests that this class of transcription factors may play a fundamental role in the development of all eukaryotic species.

1.5.2 The Homeodomain Consensus

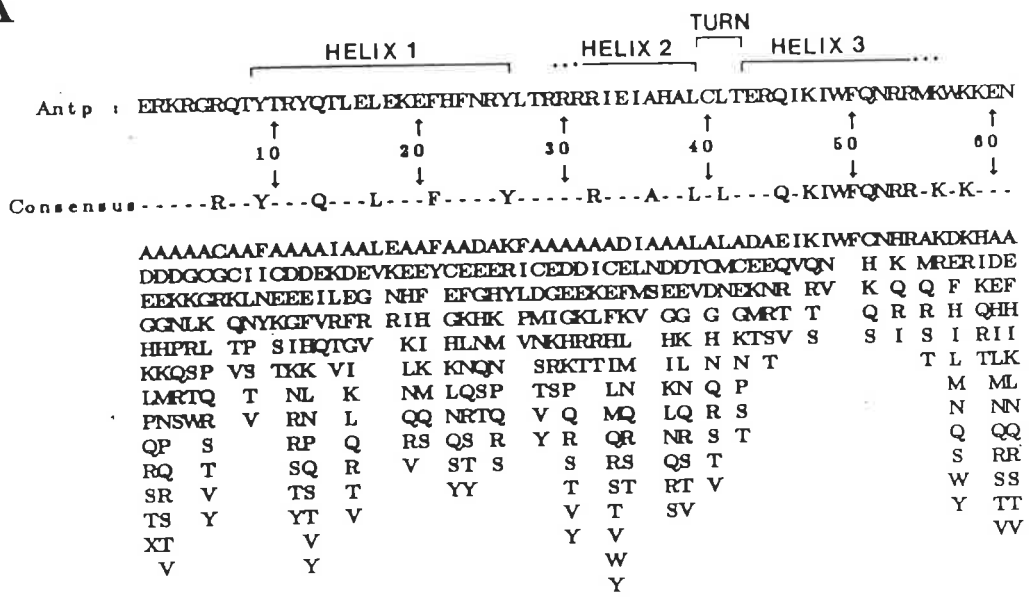
The 180 bp homeobox encodes a conserved 60 amino acid DNA binding domain, termed the homeodomain. The homeodomain motif is structurally manifest as 3 alpha-helices and a beta-turn which exhibit structural similarity to the helix-turn-helix class of DNA binding factors (Brennan and Matthews, 1989; see below for homeodomain structure). Comparison of homeodomain sequences from many eukaryotic species has revealed that at the primary sequence level there is considerable conservation of specific amino acid residues at defined positions within the homeodomain (Fig. 1.8A; Scott *et al.*, 1989; Burglin, 1994). This sequence conservation is most striking in the DNA binding residues of helix 3, the so called recognition helix, and provides a "trademark" for the identification of homeodomain proteins (Burglin, 1994). Classification of an amino acid sequence as a homeodomain

Figure 1.8 The homeodomain sequence and structure.

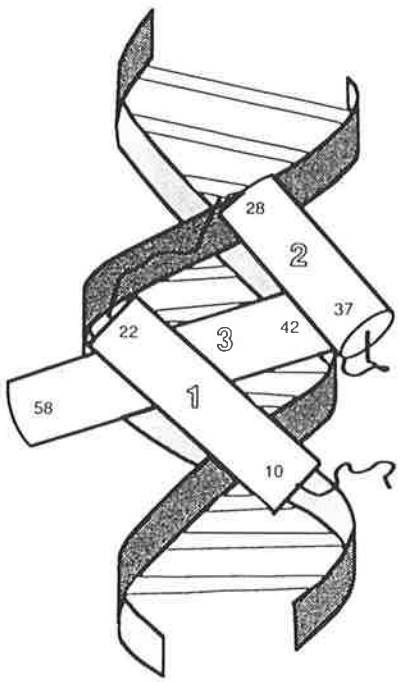
A. The homeodomain consensus sequence derived from comparison between 83 higher eukaryotic sequences is shown. Dashes indicate positions which are not conserved. Amino acid residues which have been identified at each position are shown below the consensus sequence. The *Antennapedia* sequence is shown for reference. **B.** Schematic representation of the *Engrailed* homeodomain showing the relationship between the alpha helices and the double stranded DNA. Cylinders show the position of the alpha helices. The ribbons represent the sugar-phosphate back-bone of the DNA and the bars indicate base pairs. The homeodomain residues comprising the alpha helices are also shown. **C.** Lateral view of the *Engrailed* homeodomain/DNA complex summarizing the base contacts.

A. Reproduced from Scott *et al.*, 1989. B. C. Reproduced from Kissinger *et al.*, 1990.

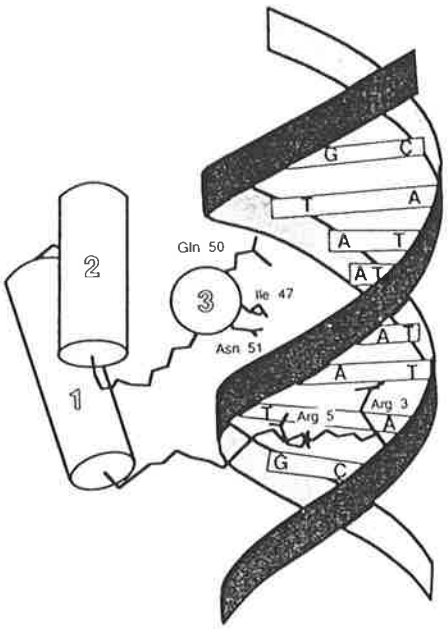
A


TURN
HELIX 1
HELIX 2
HELIX 3
Antp : ERKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN
10
20
30
40
50
60
Consensus-----R--Y---Q---L---F----Y-----R---A--L-L---Q-KIWFQNRR-K-K---

B


28
2
37
22
3
42
58
1
10

C


2
3
1
Gln 50
Ile 47
Asn 51
Arg 5
Arg 3
G
C
T
A
A
T
A
T
A
T
A
T
T
A
G
C

requires the presence of this trademark sequence, as well as most of the conserved amino acid residues which span the homeodomain.

1.5.3 3-Dimensional Structure of the Homeodomain

The 3-dimensional structure of the *Antennapedia* (modified by one residue) and *Engrailed* homeodomains bound to cognate DNA has been solved using NMR and X-ray crystallography (Otting *et al.*, 1990; Kissinger *et al.*, 1990). Although these homeodomain sequences have less than 50% identity, they are proposed to bind DNA via almost identical tertiary structures, mediated by the conserved amino acid residues which define the homeodomain. This suggests that the *Antennapedia* and *Engrailed* structures should serve as good models for the structures and DNA contacts of nearly all homeodomains.

The homeodomain structure is made up of 3 helices: helix 1 (residues 10-22) helix 2 (28-37) and helix 3 (42-58; Fig. 1.8). Homeodomain interaction with DNA occurs mostly via the highly conserved amino acid residues in helix 3, which sit in the major groove of the DNA double helix, and through an N-terminal arm which contacts the minor groove (Fig. 1.8B,C). Helices 1 and 2 pack against each other in an antiparallel arrangement and are roughly perpendicular to helix 3. The entire structure is held together by a hydrophobic core of amino acids at the interface of the three helices.

Elucidation of the *Engrailed* tertiary structure of the homeodomain/DNA complex supports the implied functional role of the highly conserved amino acid residues which define the homeodomain (Kissinger *et al.*, 1990). The invarient tryptophan and phenylalanine residues at positions 48 and 49 form part of the hydrophobic core of the homeodomain, and are essential for the correct juxtaposition of helices 1 and 2 and the N-terminal arm. Most of the remaining residues in helix 3 make base specific contacts (Asparagine-51, Isoleucine-47, Glutamine-50) and/or interact with the sugar-phosphate DNA backbone (Tryptophan-48, Arginine-53, Lysine-55, Lysine-57). Outside Helix 3, highly conserved tyrosine and arginine residues at positions 25 and 31 respectively, stabilize homeodomain binding by interaction with phosphates in the DNA backbone, and arginines at positions 3 and 5 make base specific contacts in the minor groove.

1.5.4 Homeodomain Classes

Within the broad sequence consensus, homeodomain sequences can be sub-divided into discrete classes on the basis of additional sequence conservation across the homeodomain. Individual members within a class are generally greater than 60% related at the primary sequence level (Kappen *et al.*, 1993) whereas sequence homology between classes rarely exceeds 50% (Scott *et al.*, 1989). A consensus sequence for each homeodomain class is derived from comparison of the individual homeodomain sequences which comprise that class (Scott *et al.*, 1989). The classification of homeodomain sequences into discrete classes has proved useful in determining the evolutionary relationship between different homeobox genes (Kappen *et al.*, 1993), and also provides tools (such as degenerate oligonucleotides) for the isolation of additional homeobox sequences.

Apart from direct sequence comparison, some homeobox genes are grouped on the basis of additional conserved sequence motifs contained within the homeoprotein. For example, the POU class of homeobox genes encode a highly conserved POU domain which has been shown to act in concert with the homeodomain to contribute to DNA binding specificity (Herr *et al.*, 1988; Scholer, 1991). Other examples of additional conserved motifs encoded by homeobox genes are the paired box (DNA binding; Treisman *et al.*, 1991), the LIM domain (Zn^{2+} binding; Rabbitts and Boehm, 1990), the cut repeat (unknown function; Blochlinger *et al.*, 1988; Neufeld *et al.*, 1992) and zinc fingers (DNA binding; Kaptein, 1991). Individual members within homeobox gene classes, defined by the presence of additional sequence motifs, usually encode homeodomain sequences which are also closely related to members of the same class.

1.5.5 Homeobox Genes are Transcriptional Regulators

Specific interaction between several *Drosophila* homeodomain proteins and DNA was initially established using *in vitro* binding studies (reviewed in Hayashi and Scott, 1990). This approach identified consensus sequences bound by different homeodomains *in vitro*, and laid the foundation for the identification of regulatory sequences bound by homeobox gene products *in vivo*.

The ability of homeoproteins to regulate transcription specifically in a binding site dependent manner was confirmed in cultured cell assays using *Drosophila* Schneider cells (Jaynes and O'Farrell, 1988; Han *et al.*, 1989; Krasnow *et al.*, 1989; Winslow *et al.*, 1989). Co-transfection of cells with homeobox gene expression constructs and a reporter plasmid containing the promoter of interest demonstrated that homeobox genes can activate or repress transcription (Hayashi and Scott, 1990). In some cases, the transcriptional effect depended on the target promoter. For example, *Ultrabithorax* protein activates its own promoter, but represses the high basal activity of the *Antennapedia* P1 promoter (Krasnow *et al.*, 1989). This correlates well with *in vivo* studies which show that *Ubx* positively autoregulates in the visceral mesoderm of fly embryos (Bienz and Tremml, 1988), while it negatively regulates *Antennapedia* in the central nervous system (Hafen *et al.*, 1984; Carroll *et al.*, 1986). In addition to cell culture assays, *Drosophila* homeobox gene products expressed in yeast cells are able to regulate transcription (Samson *et al.*, 1989).

Although only some of the transcriptional activities of *Drosophila* homeoproteins detected *in vitro* are consistent with *in vivo* studies, the *in vitro* approach has provided clear evidence that homeobox genes are transcription factors. The role of homeobox gene products as transcription factors *in vivo* has now been established. A good example of direct homeodomain-DNA interaction is the autoregulation of the *fushi tarazu* (*ftz*) gene in *Drosophila* (Schier and Gehring, 1992). In the early embryo, *ftz* expression is regulated by the upstream enhancer element which directs *ftz* expression in seven stripes (Pick *et al.*, 1990). *ftz*$^-$ mutant embryos do not form these stripes, indicating that functional *ftz* protein is required for expression in stripes. To determine whether *ftz* expression is autoregulated by a direct or an indirect mechanism, a second site-supression experiment was carried out. The *ftz* enhancer contains multiple *ftz* binding sites *in vitro*. Mutation of these sites from CCATTA to GGATTA (a bicoid high affinity site) prevents binding of *ftz* protein *in vitro* and strongly reduces the enhancer activity *in vivo*. This is due to the presence of a lysine residue at position 9 of helix 3 in the *ftz* homeodomain. Substitution of this lysine for a glutamine residue, as found in the bicoid protein, restores DNA binding *in vitro*. More importantly, this substitution suppressed the effect of the down-mutation and restored the seven stripes of *ftz*

expression in the early embryo. This experiment provided definitive evidence that homeoproteins are able to regulate transcription of target genes through specific interaction with DNA binding sites *in vivo*.

A recent study has identified transcriptional activation and repression of a target gene by two murine homeobox genes (Valarche *et al.*, 1993). *Phox2* and *Cux* bind to distinct sites in the upstream regulatory element of the *Ncam* (neural cell adhesion molecule) promoter via the homeodomain. Co-transfection experiments with a *Cux* expression plasmid and a reporter plasmid containing the *Ncam* promoter, showed a progressive decrease in promoter activity with increasing amounts of *Cux* plasmid. This inhibition could almost be completely reversed by cotransfecting an equivalent amount of *Phox2* expression vector. Furthermore, the ability of Cux and Phox2 to regulate *Ncam* promoter activity *in vivo* was supported by *in situ* localization data which showed that the *Cux* expression pattern included many NCAM-negative sites, and that *Phox2* expression was restricted to cells, or their progenitors, which also expressed NCAM. This investigation suggested that mammalian homeobox genes are able to regulate target genes by specific interaction with promoter regulatory sequences.

1.6 Murine Homeobox Genes

Approximately 80 homeobox genes have been isolated from the murine genome (Kappen *et al.*, 1993). These genes are divided into two large groups, based on evolutionary origin and genomic location called variously the *Hox* genes, and the orphan homeobox genes (Scott, 1992).

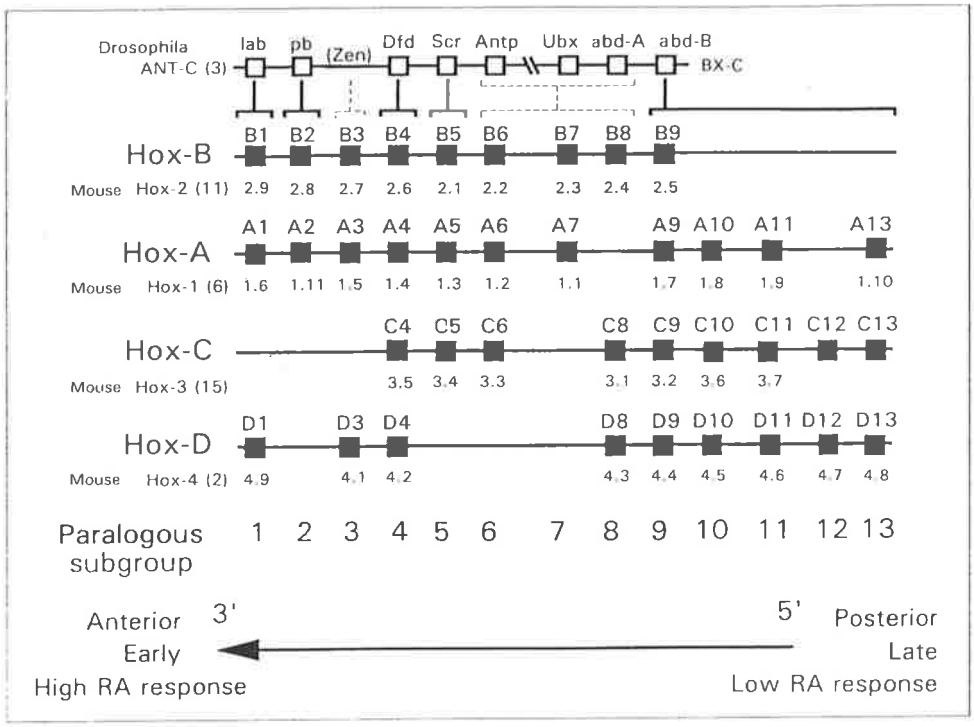
1.6.1 *Hox* Genes

The 38 murine *Hox* genes are organized into 4 related clusters each containing approximately 10 *Hox* genes, which are arranged in the same transcriptional orientation (Krumlauf, 1993a; Fig. 1.9). Based on sequence homology within the homeobox and flanking regions, the *Hox* genes can be arranged into 13 paralogous groups (Fig. 1.9). As this homology may be extended to include the homeotic genes of the Drosophila HOM-C complex (Fig. 1.9; Krumlauf, 1993a), it appears that the structure of the *Hox* clusters reflects

Figure 1.9 The *Hox* cluster.

Alignment of the four vertebrate *Hox* complexes with the *Drosophila* HOM-C homeotic complex. There are 13 paralogous groups and not all complexes have representatives of each group. The brackets indicate paralogous genes which are related to a specific *Drosophila* HOM-C homeotic gene. The letters above the boxes represent the revised nomenclature according to Scott (1992). Below the boxes are the former mouse names. The large arrow at the bottom indicates the direction of transcription and the colinear expression of the genes with respect to the anterior-posterior level, time and response to retinoic acid.

Reproduced from Krumlauf (1993a).


Drosophila
ANT-C (3)
lab
pb
(Zen)
Dfd
Scr
Antp
Ubx
abd-A
abd-B
BX-C
Hox-B
B1 B2 B3 B4 B5 B6 B7 B8 B9
Mouse Hox-2 (11) 2.9 2.8 2.7 2.6 2.1 2.2 2.3 2.4 2.5
Hox-A
A1 A2 A3 A4 A5 A6 A7 A9 A10 A11 A13
Mouse Hox-1 (6) 1.6 1.11 1.5 1.4 1.3 1.2 1.1 1.7 1.8 1.9 1.10
Hox-C
C4 C5 C6 C8 C9 C10 C11 C12 C13
Mouse Hox-3 (15) 3.5 3.4 3.3 3.1 3.2 3.6 3.7
Hox-D
D1 D3 D4 D8 D9 D10 D11 D12 D13
Mouse Hox-4 (2) 4.9 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8
Paralogous subgroup 1 2 3 4 5 6 7 8 9 10 11 12 13
Anterior 3'
Early
High RA response
5' Posterior
Late
Low RA response

their ancient origin. Molecular and phylogenetic comparisons of *Hox* gene sequence and cluster organization in a diverse range of species, suggest that the *Hox* genes arose in two stages. Firstly, tandem duplication and diversification of a single ancestral *Antennapedia*-like gene yielded a cluster of 13 genes; then the cluster was duplicated as part of a larger genomic region to yield 13 paralogous groups of *Hox* genes related by cluster duplication (Duboule and Dolle 1989; Graham *et al.*, 1989; Kappen *et al.*, 1989; Schughart *et al.*, 1989). Although the genomic organization between the 4 *Hox* clusters is similar, none of the clusters contain *Hox* gene representatives of all 13 paralogous groups. This suggests that either some cluster duplication events did not copy the entire cluster, or that some genes have been subsequently lost.

During development, the *Hox* genes are expressed in temporally and spatially restricted patterns in the paraxial mesoderm (Krumlauf, 1993b), central nervous system (Giampaolo, 1989) cranial neural crest (Wilkinson *et al.*, 1989; Hunt *et al.*, 1991), limbs (Morgan and Tabin, 1993; Izpisua-Belmonte and Duboule, 1992) and genitalia (Dolle *et al.*, 1991). A unique feature of *Hox* gene expression in these tissues is the correlation between the order of the *Hox* genes within the cluster and the temporal appearance and relative expression patterns of the genes along the embryonic axis (termed colinearity; Graham *et al.*, 1989; Duboule and Dolle, 1989; Lewis, 1978; Gaunt, 1991). For example, in the paraxial mesoderm, the 3' most gene of the *Hox* cluster has the most posterior rostrol boundary of expression, and each successive 3' gene, up to paralogous group 1, has an increasingly more anterior rostrol boundary (Fig. 1.9; Kessel and Gruss, 1990; McGinnis and Krumlauf, 1992). One exception to this rule is *Hox-b1* which has a more posterior rostrol expression boundary than *Hox-b2* (McGinnis and Krumlauf, 1992). The caudal boundary of *Hox* gene expression is less well defined, but as a general rule genes located further 5' in the cluster have successively more anterior expression boundaries (Boncinelli *et al.*, 1988; Gaunt *et al.*, 1988; Duboule and Dolle, 1989, Graham *et al.*, 1989). The temporal control of *Hox* gene expression is also reflected by the position of the *Hox* gene within the cluster (Fig. 1.9). Paralogous group 1 genes are expressed before group 2, and so on, up to paralogous group 13 (McGinnis and Krumlauf, 1992; Krumlauf, 1993a).

A further correlation between *Hox* gene position within the cluster and expression has been revealed by investigation of *Hox* gene expression during retinoic acid induced differentiation of embryonal carcinoma (EC) cells (Simeone *et al.*, 1991; Papalopulu *et al.*, 1991, Simeone *et al.*, 1992). *Hox* genes at the 5' end of the cluster are induced earlier and are in response to lower concentrations of retinoic acid than genes at the 3' end (Fig. 1.9). The induction of *Hox* gene expression during teratocarcinoma cell differentiation presumably reflects the expression of these genes in the post-gastrulation embryo, and suggests that the *Hox* genes are not expressed at high levels by pluripotent cells at early stages of development.

The widespread colinear expression of the *Hox* genes during development has led to the proposition that these genes provide a coordinated system of axial cues which form part of the mechanism for generating different positional identities within the embryo. The experimental evidence for this proposed developmental role of *Hox* genes is discussed in section 1.7.1.

1.6.2 Orphan Homeobox Genes

Apart from the four *Hox* clusters, the murine genome harbours at least 30 homeobox genes (Kappen *et al.*, 1993) termed orphan homeobox genes. The orphan homeobox genes are believed to have resulted from duplication and diversification of the *Hox* genes and encode a diverse array of homeodomain sequences. Unlike the *Hox* genes, orphan homeobox genes are not clustered, and are scattered throughout the genome.

During murine embryogenesis, orphan homeobox genes are expressed in temporally and spatially restricted patterns within a range of tissues including the brain (Simeone *et al.*, 1992; Davis and Joyner, 1988; Price *et al.*, 1991), early mesoderm (Blum *et al.*, 1992; De Jong and Meijlink, 1993), heart (Lints *et al.*, 1993), central and peripheral nervous system (Bober *et al.*, 1994; Valarche *et al.*, 1993) and spleen (Roberts *et al.*, 1994). Orphan homeobox gene expression may also be restricted to cell-types or differentiation stages rather than being spatially defined (Bedford *et al.*, 1993; Sasaki *et al.*, 1991). The diversity of orphan homeobox gene expression patterns suggest that these genes play a role in a wide

variety of developmental processes during embryogenesis. Furthermore, in many cases, orphan homeobox genes are re-expressed at different stages of development (Hill *et al.*, 1989, Davidson *et al.*, 1988; Opstelten *et al.*, 1991) indicating that a single gene may have multiple developmental functions.

1.7 The Role of Homeobox Genes in Mammalian Development

There are three lines of evidence which support a functional role of homeobox genes in mammalian development.

Firstly, molecular and genetic analysis in *Drosophila* has demonstrated that homeobox genes are key components in the establishment of pattern formation and cellular identity during fly development (reviewed in McGinnis and Krumlauf, 1992). The high conservation of homeobox gene sequence, genetic structure and expression along the anterior-posterior axis, indicates that the developmental expression and function of homeobox genes may also be conserved in vertebrates (McGinnis and Krumlauf, 1992; Duboule and Dolle, 1989; Lobe and Gruss, 1989; Kessel and Gruss, 1990). Direct support for this functional conservation has been provided by the generation of transgenic mice carrying a reporter gene construct containing the *Drosophila Deformed* (*Dfd*) autoregulatory enhancer (Awgulewitsch and Jacob, 1992). This enhancer was capable of conferring reporter gene expression to a discrete subregion of the hindbrain which correlated with the common anterior expression domain of the *Dfd* cognate *Hox* genes. This suggests that the regulatory networks which regulate *Hox* gene expression in mice, and their paralogous genes in *Drosophila*, are highly conserved.

Secondly, a developmental role for homeobox genes can be inferred from their transient and restricted expression patterns during development. As discussed above, homeobox gene expression may be restricted to specific regions of the body axes or specific regions within a variety of foetal primordia. This variety of expression patterns supports a role for homeobox genes in regional determination and cell differentiation in a range of developmental processes.

Thirdly, analyses of homeobox gene function through the generation of gain-of-function and loss-of-function mutants, has provided direct evidence for the developmental role of homeobox genes in pattern formation and determination of cellular identity.

Specific examples of homeobox gene expression and function during murine development are outlined below.

1.7.1 *Hox* Genes and Anterior-posterior Axis Pattern Formation

One of the major sites of *Hox* gene expression is in the sclerotome derivatives of the developing somites which generate the axial skeleton (reviewed in Krumlauf, 1993b; Hunt and Krumlauf, 1992). The vertebral column is made up of seven cervical vertebrae (C1-C7), thirteen thoracic vertebrae (T1-T13) which have ribs attached, six lumbar vertebrae (L1-L6) and the sacral vertebrae. The anterior expression boundary of paralogous group 1 *Hox* genes is in prevertebrae C2, and *Hox* genes with increasing paralogous number have progressively more anterior limits of expression. Several studies have shown that inappropriate *Hox* gene expression in this region during embryogenesis leads to perturbations of skeletal structures. Ectopic expression of *Hox-a7* in transgenic mice results in the homeotic transformation of the first cervical vertebrae into a posterior identity (Kessel *et al.*, 1990). Misexpression of the human *Hox-c6* (*Hox3.3*) gene in the posterior prevertebrae results in the anterior transformation of the L1 vertebrae into an anterior thoracic vertebrae (T13) (Jegalian and De Robertis, 1992). In addition, loss of function mutations for several *Hox* genes have been generated using gene knockout strategies, and the effect of these mutations on skeletal formation have been analysed. Mice homozygous for null *Hox-c8* alleles displayed a homeotic transformation of the L1 lumbar vertebra into the likeness of the last thoracic vertebra (Le Mouellic *et al.*, 1992). Mice lacking functional *Hox-b4* protein showed a homeotic transformation of the C2 vertebra to the C1 vertebra (Ramirez-Solis *et al.*, 1993). Two separate homeotic mutations were observed in *Hox-a5* mutant mice: anterior transformation of C6 into the likeness of C5, and the posterior transformation of C7 into a thoracic vertebrae (Jeannotte *et al.*, 1993).

These experiments provide direct evidence that *Hox* genes are key components in the establishment of the body plan of the mammalian embryo. Furthermore, these results are consistent with the notion that the repertoire of homeobox genes expressed by individual cells may determine, at least in part, the identity of that cell. This developmental role is presumably mediated by the ability of homeobox genes to regulate the expression of specific target genes. Additional expression and mutational analyses indicate that *Hox* genes also play a critical role in pattern formation and specification of regional identity in the cranial neural crest (Krumlauf 1993a; Hunt *et al.*, 1991; Guthrie *et al.*, 1992; Mark *et al.*, 1993;), central nervous system (Giampaolo *et al.*, 1989) and developing vertebrate limb (Dolle *et al.*, 1993; İzpisua-Belmonte *et al*, 1991).

1.7.2 *Hox11* Controls the Genesis of the Spleen

The murine orphan homeobox gene *Hox11* is expressed at 11.5 days p.c. at a spatially restricted region in the abdomen within a portion of the splanchnic mesoderm (Roberts *et al.*, 1994). Investigation of *Hox11* function during development has been carried out by generation of homozygous *Hox11* mutant mice (Roberts *et al.*, 1994). The remarkable phenotype of *Hox11* null mutant mice was the complete absence of a spleen. Morphological comparison of the splenic primordium in wild-type and null mutants indicated that the asplenia was due to a lack of cellular organization and condensation at the site where the spleen forms during embryogenesis. This experiment indicates that *Hox11* has a fundamental role in the establishment of regional identity at the initial stages of spleen development. Furthermore, at the cellular level, *Hox11* appears to be a critical component in the molecular mechanisms which determine cellular identity within the splanchnic mesoderm. Presumably, this developmental function is mediated by transcriptional regulation of downstream target genes, through specific interaction via the *Hox11* homeodomain.

1.7.3 The Role of *Goosecoid* in Embryonic Induction

Functional analysis in *Xenopus* and murine embryos have shown that the *goosecoid* homeobox gene plays a central role in mediating the formation of differentiated tissues during

gastrulation. Expression of *goosecoid* in the *Xenopus* gastrulating embryo is restricted to the dorsal blastopore lip (Cho *et al.*, 1991). Experiments have shown that this tissue, the Spemann's organizer, harbours developmental information capable of inducing the development of the entire body axis (Spemann and Mangold, 1924). Cho *et al.* (1991) showed that microinjection of *Xenopus goosecoid* mRNA into the ventral half of the embryo resulted in the generation of organizer function as seen by induction of an entire secondary axis. This demonstrated that the *goosecoid* homeobox gene was not just a marker of embryonic position, but was also capable of inducing the necessary cellular responses required for body axis formation. Therefore, at the cellular level, it appears that goosecoid has the ability, at least in part, to determine the pathway of cellular differentiation during the initial differentiation stages of *Xenopus* development.

A murine homolog of goosecoid has also been isolated (Blum *et al.*, 1992). Murine *goosecoid* is expressed during gastrulation in the anterior tip of the primitive streak (Blum *et al.*, 1992), which is the morphological equivalent of the *Xenopus* blastopore lip. Xenograft transplantation experiments of murine primitive streak tissue into *Xenopus* host gastrulae, showed that regions of murine tissue which correlated with *goosecoid* expression induced secondary axis formation (Blum *et al.*, 1992). Therefore, experimental evidence supports a conserved developmental role for *goosecoid* during murine embryogenesis. Apart from the implied functional conservation, *goosecoid* provides a good example of the ability of homeobox genes to control cellular differentiation, presumably by the specific transcriptional regulation of downstream target genes.

1.7.4 Homeobox Genes and Cellular Transformation

Cellular transformation requires the uncoupling of the molecular mechanisms which coordinate cellular proliferation and differentiation, and results in the uncontrolled expansion of cellular differentiation intermediates. Inappropriate expression of genes which determine cellular identity, therefore, has the potential to inhibit cellular differentiation and result in cellular transformation. Several examples have implicated homeobox genes in the generation of B- and T-cell leukaemias as a result of ectopic expression in haematopoietic cell lineages

(Perkins and Cory, 1993; Perkins *et al.*, 1990; Ben-David *et al.*, 1991; Kennedy *et al.*, 1991; Hatano *et al.*, 1991; Lu *et al.*, 1991; Kamps *et al.*, 1990). The established oncogenic potential of homeobox genes suggests that these genes are important for the maintenance of cellular identity when expressed within the appropriate cellular context. Therefore, the identification of homeobox genes as oncogenes provides further support for a developmental role for homeobox genes in the determination of cellular identity during embryogenesis.

1.8 Homeobox Gene Expression in Pluripotent Cells during Early Development

While homeobox gene expression in the developing tissues of post-gastrulation embryos has been the subject of extensive investigation, comparatively little is known about homeobox gene expression in the pluripotent cells during early embryogenesis. Two POU class homeobox genes which are expressed in embryonic lineages of the early embryo are *Oct-4* (Scholer *et al.*, 1990a; Rosner *et al.*, 1990; Yeom *et al.*, 1991; Okamoto *et al.*, 1990) and *Oct-6* (Suzuki *et al.*, 1990; Meijer *et al.*, 1990). During embryogenesis, *Oct-4* is expressed by the totipotent and pluripotent cell lineages which include the fertilized egg, morula, inner cell mass, primitive ectoderm, primordial germ cells, testis and ovary, and is downregulated during differentiation of these cells (Rosner *et al.*, 1990; Yeom *et al.*, 1991; Scholer *et al.*, 1990b). *Oct-4* therefore provides a useful marker for the identification of pluripotent cells (Rosner *et al.*, 1991). Like *Oct-4*, *Oct-6* is also expressed in the inner cell mass of the blastocyst (Scholer, 1991) but has also been detected later in embryogenesis in specific neurons of the developing brain (Suzuki *et al.*, 1990). Both *Oct-4* and *Oct-6* are expressed in the pluripotent ES and embryonal carcinoma (EC) cells (Scholer, 1991), and are downregulated when the cells are induced to differentiate with chemical inducers (Suzuki *et al.*, 1990; Scholer *et al.*, 1990b; Okamoto *et al.*, 1990; Rosner *et al.*, 1990; Scholer *et al.*, 1990a). This expression pattern *in vitro* presumably reflects the downregulation of *Oct-4* and *Oct-6* during pluripotent cell differentiation *in vivo*. The decrease in *Oct-4* and *Oct-6* expression during differentiation *in vitro*, and loss of pluripotency *in vivo*, indicate that these genes, particularly *Oct-4*, may be required to maintain pluripotency during early

embryogenesis. However, functional analysis is required to determine the precise mechanism by which these homeobox genes carry out this proposed developmental function.

1.9 Aims and Approaches

Embryonic stem (ES) cells are derived from the pluripotent cells of the early murine embryo, and can be maintained *in vitro* in the undifferentiated state, or induced to differentiate by withdrawal of the cytokine LIF, or by exposure to chemical inducers. ES cells reintroduced into the blastocyst, are able to colonize all embryonic tissues including the germ line, indicating that ES cells cultured *in vitro* maintain pluripotency. The ES cell culture system therefore allows direct investigation of developmental gene function during pluripotent cell differentiation and proliferation both *in vitro* and *in vivo*. Furthermore, as ES cells cultured *in vitro* maintain pluripotency, these cells are a readily available resource for the isolation of developmental control genes which are expressed by the pluripotent progenitor cells of the early murine embryo.

Homeobox genes play a key role in developmental processes such as pattern formation and the specification of regional and cellular identity in the post-gastrulation embryo. Formation of the broad spectrum of tissue and cell types which constitute the post-gastrulation embryo results directly from the controlled expansion and directed differentiation of the pluripotent cells of the pre-implantation embryo. It likely, therefore, that homeobox genes expressed by pluripotent cells play a fundamental developmental role during murine development by specification of pluripotent cell identity, and control of pluripotent cell differentiation and proliferation.

The broad aim of this thesis was to isolate homeobox genes which play a developmental role in pluripotent cell differentiation and proliferation during early development. As ES cells provide a source for the isolation of genes expressed by pluripotent cells, the initial approach was to determine the range of homeobox genes expressed by undifferentiated ES cells. Further analysis of homeobox gene expression in a range of differentiated ES cells induced by chemical and spontaneous induction was then carried out,

to provide some insight into the expression, and perhaps developmental role, of these genes during early murine development.

The developmental role of homeobox genes in the specification of regional and cellular identity, is reflected by the temporal and spatial restriction of homeobox gene expression during embryogenesis. This relationship between homeobox gene expression and cellular identity, indicates that homeobox genes can be exploited as molecular markers for the identification of specific cell types. Indeed, one example of this feature of homeobox genes is the restricted expression of *Oct-4* to the pluripotent cell lineages in the murine embryo (Rosner *et al.*, 1991; see 1.8).

Effective analysis of early murine development *in vitro* requires cell-type specific markers which allow correlation between the distinct cell types and differentiation pathways generated *in vitro* with their *in vivo* equivalents. Few markers for pluripotent cells have been isolated, and are basically restricted to *Oct-4* (Scholer *et al.*, 1990a; Rosner *et al.*, 1990; Yeom *et al.*, 1991; Okamoto *et al.*, 1990), alkaline phosphatase (Adamson, 1988) and SSEA-1 (Solter and Knowles, 1978). Therefore, a further aim of this thesis was to isolate homeobox genes which could be used as ES cell specific markers. As the pluripotent cell population within the early embryo which gives rise to ES cells has not been rigourously defined, the identification of ES cell specific markers will aid in determining the origin of ES cells. Furthermore, pluripotent cell markers will aid in the identification and isolation of commercially valuable ES cells from domestic and livestock species.

CHAPTER 2: MATERIALS AND METHODS

CHAPTER TWO: MATERIALS AND METHODS

2.1 Abbreviations

Abbreviations are as described in "Instructions to authors" (1978) *Biochem. J.* **169**, 1-27.

In addition:

A_{xxx}	absorbance at xxx nm
bp	base pairs
BCIG	5-bromo-4-chloro-3-indolyl-beta-D-galactoside
BCIP	5-bromo-4-chloro-3-indolyl-phosphate
bisacrylamide	N, N'-methylene-bisacrylamide
BSA	Bovine serum albumin (Fraction V)
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DTT	dithiothreitol
EtBr	Ethidium Bromide
FCS	Foetal calf serum
IPTG	Isopropylthiogalactoside
kb	kilobase pairs
MBA	3-methoxybenzamide
MOPS	3-[N-Morpholino]propane-sulfonic acid
NBT	nitro blue tetrazolium chloride
PBS	phosphate buffered saline
PEG	polyethylene glycol
RA	*all-trans*-Retinoic acid
RNAase	ribonuclease
SDS	sodium dodecyl phosphate
TEMED	N,N,N',N'-tetramethyl-ethenediamine

2.2 Materials

2.2.1 Chemicals and Reagents

Sigma Chemical Co. supplied the following chemicals: Acrylamide, bisacrylamide , agarose (Type 1), ampicillin, BSA, EtBr, ethylenediaminetetra-acetic acid (EDTA), Heparin, MOPS, RA, ribonucleotide phosphates (rNTPs), salmon sperm DNA, SDS, TEMED, Tris base, tRNA (from brewer's yeast).

Sources for other important reagents were as follows: MBA; Aldrich. IPTG, BCIG; Progen. BCIP, Anti-DIG Fab antibody-alkaline phosphatase conjugate, NBT and glycogen; Boehringer Mannheim. Ammonium persulphate, phenol and PEG 6000; BDH Chemicals. Paraformaldehyde; Merck. Poly dT_{18} primer; New England Biolabs. Sepharose CL-6B; Pharmacia. Gluteraldehyde; Probing and Structure. DTT; Diagnostic Chemicals Ltd.

All chemicals and reagents were of analytical grade.

2.2.2 Radiochemicals

[alpha-^{32}P] dATP (3000 Ci/mmol), [alpha-^{35}S] dATP (1000-1500 Ci/mmol) and [alpha-^{32}P] UTP (3000 Ci/mmol) were supplied by Bresatec.

2.2.3 Kits

^{T7}Sequenase kit:	Pharmacia
Gigaprime kit:	Bresatec
The Geneclean II™:	Bresatec
Alkaline Phosphatase staining kit:	Sigma Diagnostics
Giemsa staining kit:	Sigma

2.2.4 Enzymes

Restriction endonucleases were supplied by Pharmacia and New England Biolabs. Other enzymes were obtained from the following sources: Calf intestinal phosphatase (CIP) and Proteinase K; Boehringer Mannheim. *E. coli* DNA polymerase I (Klenow fragment), T_{aq}

Polymerase, RNAasin, DNAase 1 and T4 DNA ligase, T7 RNA polymerase; Bresatec. Lysozyme and Ribonuclease A (RNAase A); Sigma. Ribonuclease T1 (RNAase T1); Ambion. T7 DNA polymerase; Pharmacia. Avian Myeloblastosis Virus (AMV) reverse transcriptase; Molecular Genetic Resources. T3 RNA polymerase; Promega.

2.2.5 Buffers

General Buffers:

Denhardt's solution: 0.1% (w/v) Ficoll, 0.1% (w/v) polyvinylpyrrolidone, 0.1% (w/v) BSA

SSC: 150 mM NaCl, 15 mM sodium citrate

TBE: 90 mM Tris, 90 mM boric acid, 2.5 mM EDTA, pH 8.3

GTE: 50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA, pH 7.6

GEB: 10 mM Tris-HCl, 1 mM EDTA, 0.1% SDS, pH 7.5

FLB: 95% (w/v) deionised formamide, 0.02% bromophenol blue, 0.02% xylene cyanol

GLB: 50% glycerol, 0.1% SDS, 0.05% bromophenol blue, 0.05% xylene cyanol

MOPS: 23 mM MOPS pH 7.0, 50 mM NaAcetate, 10 mM EDTA

TNM: 30 mM Tris-HCl pH 7.6, 150 mM NaCl, 15 mM $MgCl_2$, 0.4% NP40

TUNES: 10 mM Tris-HCl pH 8.0, 7 M urea, 0.35 M NaCl, 1 mM EDTA, 2% SDS

RIPA: 150 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl pH 8.0, 0.5% sodium deoxycholate, 0.1% SDS

STET: 50 mM Tris-HCl pH 8.0, 50 mM EDTA, 8% sucrose, 5% Triton X100

TBST: 136 mM NaCl, 2.7 mM KCl, 25 mM Tris-HCl pH 7.5, 0.1% Tween 20

AP buffer 100 mM NaCl, 50 mM $MgCl_2$, 100 mM Tris-HCl pH 9.5, 0.1% Tween 20

BBS: 50 mM BES pH 6.99, 280 mM NaCl, 1.5 mM Na_2HPO_4

All general buffers were sterilized by autoclaving.

DIG riboprobe dot blot buffers

Buffer 1: 100 mM Tris-HCl pH 7.5, 150 mM NaCl

Buffer 2: 0.5g blocking reagent (Bresatec) in 10 ml Buffer 1

Buffer 3: 100 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$

Buffer 4: 10 mM Tris-HCl pH 8.0, 1 mM EDTA

DIG riboprobe dot bot buffers were prepared from autoclaved solutions.

2.2.6 Cloning and Expression Vectors

pT7T319U was a kind gift from Dr. Helena Richardson.

The p1OX and p3OX expression vectors were kindly supplied by Ms. Linda Whyatt.

2.2.7 Cloned DNA sequences

The *Oct-4* cDNA clone in pBluescript was a kind gift from Dr. Hans Scholer. This clone contained a 462 bp *Stu* I cDNA fragment spanning positions 491 to 953 of the *Oct-4* cDNA sequence (Scholer *et al.*, 1990).

The pPGKNeo construct (Whyatt *et al.*, 1993) was kindly supplied by Ms. Linda Whyatt.

2.2.8 Oligonucleotides

Synthetic DNA primers were synthesized by Bresatec. All primer sequences are 5'-3' (N = A, C, G or T; / indicates positions of degeneracy).

General Sequencing Primers:

T7: TAATACGACTCACTATAGGGAGA

T3: ATTAACCCTCACTAAAGGGA

m13: GTAAAACGACGGCCAGT

Hesx1 Sequencing Primers:

rep5: TAGAATTCTAGAACAGTACA

rep3: TAGAATTCACAAATAGCTCA

Homeobox PCR primers:

Ant5: GAATTCGAA/GC/TTCGAA/GAAA/GGAA/GTTC/TCAT

Ant3: AAGCTTC/TTCC/TTTC/TTTA/GTGC/TTTCATA/G/TCG

En5: AAA/GA/CGNCCAA/CGNACNGCA/CTT

En3: TTA/GTTC/TTGA/GAACCAGATC/TTTA/G/TAT

5'mANT: TAGAATTCGAA/GC/TTNGAA/GAAA/GGAA/GTT

5'mEN: TAGAATTCA/CGNCCNA/CGNACNGCNTT

3'ALL: TAGAATTCCG/TNCG/TA/GTTC/TTGA/GAACCA

Hesx1 ORF primers (used in *Waved coat* analysis)

Wc 1/5: TAGAATTCAGCTCTGCACACGT

Wc 1/3: TAGAATTCACAAGTTGAGTTGCA

Wc 2/5: TAGAATTCTGTAAGATGGACCTT

Wc 2/3: TAGAATTCTTGTCACTATTGGCT

Wc 3/5: TAGAATTCCTTCTTATT

Wc 3/3: TAGAATTCTGTACTATACAA

Miscellaneous Primers

3'Hesx1a*: TAGAATTCGAAGCCAACATGTGT

2.2.9 Bacterial Strains

The *E. coli* DH5alpha strain (a kind gift from Dr. Helena Richardson) was used as a host for all recombinant plasmids. Plating bacteria for library screening were prepared from BB4 strain *E. coli*, and cDNA library clones were zapped using the XL1-Blue strain (these strains were purchased from Clontech). *E. coli* strain phenotypes were as follows:

DH5alpha: *supE*44 delta*lac* U169 (phi80 *lacZ*deltaM15) *hsdR*17 *recA*1 *endA*1*gyrA*96 *thi*-1*relA*1

BB4: *supF*58 *supE*44 *hsdR*514 *galK*2 *galT*22 *trpR*55 *metB*1 *tonA* delta*lac*U169 F'[*proAB*$^+$ *lacI*^{q} *lacZ*deltaM15 Tn*10*(*tet*^{r})]

XL1-Blue: *supE*44 *hsdR*17 *recA*1 *endA*1 *gyrA*46 *thi relA*1 *lac*$^-$ F'[*proAB*$^+$ *lacI*^{q} *lacZ*deltaM15 Tn*10*(*tet*^{r})]

Stock cultures of these strains (and transformed bacteria) were prepared by dilution of an overnight culture with an equal volume of 80% glycerol and storage at -80°C.

2.2.10 Bacterial Growth Media

Growth media were prepared in double distiled water and sterilised by autoclaving. Antibiotics and other labile chemicals were added after the media solution had cooled to 50°C.

Luria (L) broth: 1% (w/v) Bacto-tryptone (Difco), 0.5% (w/v) yeast extract (Difco), 1% (w/v) NaCl, adjusted to pH 7.0 with NaOH.

LMM broth: 1% (w/v) Bacto-tryptone (Difco), 0.5% (w/v) NaCl, 0.4% maltose, 0.2% $MgSO_4$

Solid Media: Agar plates were prepared by supplementing the above media with 1.5% Bacto-agar (Difco). LMM agarose for library screening was prepared by dissolving 0.7 g agarose/100 ml LMM broth.

Ampicillin (100 µg/ml) was added where appropriate for growth of transformed bacteria to maintain selective pressure for recombinant plasmids.

2.2.11 Tissue Culture Cell Lines and Media

2.2.11.1 ES Cell Lines

The sources of the following ES cell lines used throughout the course of this work were: MBL-5 and D3 (Lindsay Williams, Ludwig Institute, Melbourne, Australia), E14 (Anna

Michelska, Murdock Institute, Melbourne, Australia), E14TG2A (Austin Smith, CGR Edinburgh, UK), J1 (Lyn Corchrine, Walter and Eliza Hall Institute, Melbourne, Australia), CGR8 (Austin Smith, CGR Eninburgh, UK), CCE (Richard Harvey, Walter and Eliza Hall Institute, Melbourne, Australia).

2.2.11.2 Other Cell Lines

MED2 cells (unpublished data).

2.2.11.3 Solutions

Phosphate buffered saline (PBS): 136 mM NaCl, 2.6 mM KCl, 1.5 mM KH_2PO_4, 8mM Na_2HPO_4, pH 7.4. Sterilized by autoclaving (20 psi for 25 minutes at 140°C).

Trypsin: 0.1% trypsin (Difco) and 1 X EDTA Versene buffer solution (CSL). Sterilized by filtration through a 0.2 µM filter (Whatman).

Beta-mercaptoethanol/PBS: 100 mM Beta-mercaptoethanol (Sigma) in 14 ml PBS. Fresh solution was made every two weeks.

2.2.11.4 Media

ES DMEM medium: 67.4g Dulbecco's Modified Eagle Medium (Gibco BRL), 18.5g $NaHCO_3$ and 6.25 ml (40 mg/ml) Gentamicin (Delta West) dissolved in 5 litres sterile water.

Hams F12 medium: 1 packet F12 Nutrient mix (Gibco BRL), 5.88g $NaHCO_3$, 6.25 ml (40 mg/ml) Gentamicin (Delta West) in 5 litres sterile water.

ES cell selection medium: Complete ES medium with 200 µg/ml Gentamicin (G418-Sulfate, Gibco BRL).

Incomplete ES cell medium: 90% ES DMEM medium, 10% FCS (Commonwealth Serum Laboratories), 1% glutamine, 0.1% Beta-mercaptoethanol/PBS.

Complete ES cell medium: Incomplete ES cell medium with 1% COS cell conditioned medium containing LIF (Smith, 1991).

X cell medium: 1:1 MED2 conditioned medium:ES cell medium (complete or incomplete)

MED2 conditioned medium: media was isolated from MED2 cells cultured in incomplete ES cell medium for 5 days.

All tissue culture media were filter sterilized.

2.2.12 DNA Markers

Hpa II digested pUC19 markers were purchased from Bresatec Ltd. Band sizes (bp): 501, 489, 404, 331, 242, 190, 147, 111, 110, 67, 34, 26.

Eco R1 / *Hind* III lambda DNA markers were prepared by digestion of lambda DNA (New England Biolabs). Band sizes (bp): 21227, 5148, 4973, 4268, 3530, 2027, 1904, 1584, 1375, 974, 831, 564, 125.

2.2.13 Miscellaneous Materials

3MM chromatography paper: Whatman Ltd.

X-Omat AR diagnostic film: Kodak

Medical grade X-ray film: Konica

Nytran nylon membrane (0.45 μM): Schleicher and Schuell

petri-dishes, 6-, 24-, 96-well trays: Corning, Falcon

2.3 Molecular Methods

2.3.1 Restriction Endonuclease Digestion of DNA

Plasmid DNA was digested with 1-2 units of enzyme/μg DNA and was incubated at the appropriate temperature for 1-5 hours. Genomic DNA was digested overnight with 5 units of enzyme/μg DNA. Appropriate reaction buffers were determined using the NEBuffer system (New England Biolabs). Plasmid and genomic DNA was assayed for complete digestion by agarose gel electrophoresis.

2.3.2 Gel Electrophoresis

Agarose gel electrophoresis

Agarose gel electrophoresis (1% to 2% w/v agarose in 1 X TBE, stored at 68°C) was carried out using horizontal mini-gels prepared by pouring 10 ml of gel solution onto a 7.5 cm x 5.0 cm glass microscope slide. Agarose mini-gels were submerged in 1 X TBE and samples containing 1 X GLB were electrophoresed at 80 mA. Nucleic acid was visualised by staining gels with EtBr (4 mg/ml in water) and exposure to medium wavelength UV light. Appropriate bands were removed from preparative gels using a sterile scalpel blade.

Polyacrylamide gel electrophoresis

Non-denaturing gels

A 30% acrylamide stock solution was prepared by dissolving 29g acrylamide and 1g bis-acrylamide in a final volume of 100 ml MQ water. Polyacrylamide gels were prepared by pouring a 50 ml gel solution containing 5%-10% acrylamide in 1 X TBE between 15 cm x 16 cm glass plates separated by 1 mm spacers. Prior to pouring the gel, 500 µl of freshly prepared 10% ammonium persulphate and 60 µl TEMED were added to the gel solution to initiate polymerisation. The gel was allowed to set for 30-60 minutes and pre-electrophoresed at 25 mA in 1 X TBE for 10 minutes. DNA samples containing 1 X GLB were electrophoresed at 25 mA. After electrophoreses, the glass plates were separated and the gel stained with EtBr, destained in water for 10 minutes, and DNA bands excised under medium wavelength UV light using a sterile scalpel blade.

Denaturing gels

6% polyacrylamide sequencing gels were prepared from a 40% acrylamide stock solution (38:2, acrylamide:bis-acrylamide). All other denaturing acrylamide gels were prepared from a 30% acrylamide solution (29:1, acrylamide:bis-acrylamide). Denaturing polyacrylamide gels were prepared by dissolving 21 g urea (final concentration 7 M) in sterile distiled water with subsequent addition of 2.5 ml 20 X TBE and an appropriate volume of acrylamide stock solution in a final volume of 50 ml. 500 µl of freshly prepared 10% ammonium persulphate and 60 µl of TEMED were added to the gel solution prior to

pouring between clean glass plates (20 X 40 cm) separated by 0.4 mm spacers. Once the gel had set (approximately 30 minutes), the comb was removed and wells flushed with water. Denaturing gels were pre-electrophoresed for 30 minutes at 1400 V and the wells were flushed with 1 X TBE before loading samples. Gels were electrophoresed at 1400 V (about 50°C). After the gel had run, the glass plates were prised apart and the gel briefly fixed with 20% ethanol / 10% acetic acid before transferring to wet 3MM Whatmann paper. The excess liquid was removed and the gel dried down at 65°C under vacuum. Radioactivity was detected by exposure to X-ray film at -80°C with intensifying screens, or by exposure to phosorimager screens.

2.3.3 Purification of Linear DNA Fragments

Linear DNA fragments greater than 500 bp were purified using the GenecleanII™ kit. Smaller DNA fragments were separated by non-denaturing gel electrophoresis (2.3.2), excised and eluted overnight in 400 µl GEB. The following day, the eluate was transferred to a fresh eppendorf tube and the DNA was precipitated by addition of 40 µl 3 M NaAc, 800 µl ethanol and 1 µl glycogen and chilled at -80°C for 30 minutes. After centrifugation at 14 K for 15 minutes at room temperature the DNA was resuspended in MQ water.

2.3.4 DNA Gels for Southern Blot

Agarose gels for southern blot analysis were prepared as described in section 2.3.2. Prior to transfer, southern blot gels were soaked in 0.25 M HCl for 15 minutes (depurnation), 1.5 M NaCl / 0.5 M NaOH for 2 X 15 minutes (denatuation), and 3 M NaAc pH 5.5 for 30 minutes (neutralization). DNA was transferred to Nytran Nylon membrane as described in 2.3.20.

2.3.5 DNA Ligation Reactions

Ligation reactions contained 25 ng purified vector, 50-100 ng DNA insert, ligation buffer (50 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP) and 1 unit T4 DNA ligase, and were incubated at room temperature for 1-5 hours.

2.3.6 Preparation of Competent Cells

Overnight cultures (50 ml, L broth) were subcultured by transferring 500 µl into 50 ml of fresh L broth and incubating in an orbital shaker at 37°C for approximately 2 hours. Once the culture was at OD = 0.4, the cells were placed on ice for 15 minutes and centrifuged at 4K for 5 minutes at 4°C. The bacterial pellet was gently resuspended in 0.1 M $MgCl_2$ (5 ml) and placed on ice for 15 minutes. The cells were then centrifuged at 3K for 10 minutes at 0°C, the pellet gently resuspended in cold 100 mM $CaCl_2$ (2 ml). Competent cells were stored at 4°C for at least 60 minutes before transformation.

2.3.7 Transformation of Competent Cells

Approximately 10 ng ligated DNA (2.3.5) was added to 200 ml of competent cells and placed on ice for 30 minutes. The cells were heat shocked for 2 minutes at 42°C before cooling on ice for 30 minutes. After incubation at room temperature for 10 minutes, 500 ml of sterile L broth was added, and the cells incubated at 37°C for 20 minutes on a rotating drum. L broth ampicillin (100 mg/ml) plates were spread with 200 ml of culture and incubated overnight at 37°C.

2.3.8 Rapid Small Scale Preparation of DNA (Mini-prep)

1.5 ml L broth (ampicillin 100 µg/ml) was inoculated with a single transformant colony and incubated overnight at 37°C with shaking. Cell cultures were transferred to eppendorf tubes, spun for 1 minute, and bacterial pellets thoroughly resuspended in 200 µl STET buffer. 10 µl lysozyme (10 mg/ml) was added and the bacteria lysed at 100°C for 45 seconds. Samples were centrifuged for 15 minutes and the cell debris and chromosomal DNA removed with a sterile toothpick. Plasmid DNA was precipitated by addition of 200 µl isopropanol and incubation on ice for 5 minutes. Plasmid DNA was centrifuged for 10 minutes, washed with 70% ethanol, and resuspended in 20 µl MQ H_2O.

2.3.9 Double Stranded Sequencing of Plasmid DNA

Denaturation of plasmid DNA

10 µl mini-prep DNA (2.3.8) was diluted to 20 µl with MQ H_20, and 1 µl RNAase A (20 mg/ml) was added before incubation at 37°C for 15 minutes. Plasmid DNA was denatured with by the addition of 5 µl 1 M NaOH / 1 mM EDTA and placed at 37°C for 15 minutes. Denatured plasmid was purified by centrifugation at 1.8 K for 3 minutes on a Sepharose CL-6B column.

Sequencing of plasmid DNA

Dideoxy sequencing reactions were carried out using the Pharmacia ^{T7}polymerase sequencing kit according to manufacturer's instructions. Reaction products were separated on a 6% polyacrylamide denaturing gel.

2.3.10 Library Screening

(i) Preparation of plating bacteria

50 ml LMM broth was inoculated with a single bacterial colony and grown overnight at 37°C in an orbital shaker. Bacterial cultures were centrifuged at 5 K for 5 minutes at 4°C and resuspended in 20 ml 10 mM $MgSO_4$.

(ii) Library plating

5 X 10^4 recombinant phage were added to 200 µl plating bacteria and incubated at 37°C for 20 minutes. 8 ml LMM agarose at 42°C was transferred to the phage / bacteria mixture, briefly mixed and overlayed onto 15 cm L agar plates. Bacteriophage plaques were formed by incubating at 37°C for 4-5 hours. When the plaques had grown to a sufficient size, the plates were transferred to 4°C for at least 1 hour before plaque lifts.

(iii) Plaque lifts

Nitrocellulose circles were placed onto agar plates for 90 seconds (1st lift) 5 minutes (second lift) and transferred to 3MM Whatmann paper soaked in denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 5 minutes. Filters were then transferred to 3MM Whatmann paper soaked in neutralising solution (0.5 M Tris-HCl pH 8.0, 1.5 M NaCl) for 5 minutes and

rinsed in 20 X SSC for 5 minutes. After air drying filters for 30 minutes, nucleic acid was bound to the filter by UV crosslinking.

2.3.11 Lambda Zapping

cDNA clones were isolated from third round positive plaques using the zapping protocol provided by Clontech.

2.3.12 Large Scale Plasmid Preparation

250 ml Lbroth + ampicillin (100 µg/ml) was inoculated either with a single bacterial colony or from a glycerol stock, and incubated overnight at 37°C in an orbital shaker. The cells were harvested by centrifugation at 7 K for 5 minutes at 4°C, and the bacterial pellets were briefly drained. The bacteria were thoroughly resuspended in 6.5 ml GTE before the addition of 13 ml of lysis solution (0.2 M NaOH, 1% SDS). After gently mixing on ice for 5 minutes, 6.5 ml of cold acetate solution (3 M KAc / 2 M HOAc, pH 5.8) was added and the solution was mixed well before incubating on ice for 10 minutes. The bacterial debris were pelleted (12 K / 10 minutes / 4°C), and the supernatant transferred to a new tube before precipitation with 15 ml isopropanol. The plasmid DNA was recovered by centrifugation (8 K / 5 minutes / 4°C) and gently resuspended in a solution containing 7 ml TE with 7g CsCl. Once the DNA pellet was thoroughly resuspended, 700 µl EtBr (10 mg/ml) was added, and any remaining debris removed by centrifugation at 3 K for 10 minutes at room temperature. The supernatant was transferred to a 10 ml Nalgene Oakridge tube and sealed with parafin oil. After centrifugation at 45 K for 24 hours at 20°C, the super-coiled plasmid DNA was visualised under long wavelength UV light and recovered using a 1 ml syringe. The EtBr was removed by repeated extraction with NaCl/TE saturated isopropanol. The DNA solution was diluted with 3 ml of MQ water before precipitation with 2 volumes of ethanol and incubation overnight at -20°C. Plasmid DNA was recovered by centrifugation (9 K / 30 minutes / 4°C) and resuspended in 400 µl MQ water. The DNA solution was transferred to an eppendorf tube and reprecipitated before final resuspension in MQ water. Plasmid yields were calculated from the A_{260} value determined by spectrophotometric analysis.

2.3.13 Preparation of Genomic DNA

Genomic DNA was isolated from ES cell pellets stored at -80°C. Cells were resuspended in 700 µl Tris/saline (25mM Tris 7.6,75 mM NaCl, 24 mM EDTA pH 8.0) before lysis with 70 µl 10% SDS and gentle inversion. Cellular protein was degraded by the addition of 8 µl Proteinase K (25 mg/ml) and incubation at 37°C for 2.5 hours with occasional mixing. Genomic DNA samples were extracted twice with Tris-saturated phenol and twice with chloroform before precipitation with 1/10th volume 2M KCl and 2 volumes ice-cold 100% ethanol. Genomic DNA was spooled out and transferred into 1 ml MQ H_2O and resuspended over-night.

2.3.14 Reverse Transcription

10 µg cytoplasmic RNA in a volume of 6.5 µl was denatured at 65°C for 15 minutes before snap cooling on ice. The denatured RNA was added to a reverse transcription reaction containing 50 mM Tris-HCl pH 8.5, 6 mM $MgCl_2$, 40 mM KCl, 1 mM DTT, 1.5 mM each dNTP, 40 units RNAasin, 500 ng oligo d(T) primer and 8 units AMV reverse transcriptase and incubated at 41°C for 2 hours. After incubation, the cDNA solution was diluted to 500 µl with MQ H_2O.

2.3.15 Polymerase Chain Reaction

(i) PCR with cDNA template

PCR reactions contained 5 µl cDNA, 10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% gelatin, 200 µM each dNTP, 1 µg each degenerate primer and 1 unit T_{aq} polymerase in a final volume of 20 µl.

(ii) PCR with genomic DNA template

Genomic DNA for PCR was denatured by boiling for 5 minutes and snap cooled on ice. PCR reactions contained 300 ng denatured genomic DNA, 10 mM Tris-HCl pH 8.3, 50 mM KCl, 2.0 mM $MgCl_2$, 0.001% gelatin, 200 µM each dNTP, 750 ng each primer and 1 unit T_{aq} polymerase in a final volume of 20 µl.

A typical thermal profile for PCR was as follows:

Denaturation	94°C	2 minutes	
Amplification	94°C	10 seconds	
	50°C	10 seconds	
	72°C	1 minute	35 cycles
Final Extension	72°C	2 minutes	

PCR reactions were amplified in a Corbett capillary thermal cycler.

2.3.16 Isolation of Cytoplasmic RNA from Cultured Cells

Cytoplasmic RNA was isolated using the method of Edwards *et al.* (1985). Cells were harvested by trypsinization and stored at -80°C until use. Cell pellets were thoroughly resuspended in 2 ml ice cold and TNM and lysed by vigorous pipetting 10 times. After incubation on ice for 5 minutes, nuclei were pelleted by centrifugation (3 K for 5 minutes) and the supernatant decanted and mixed thoroughly with 2 ml TUNES. This solution was extracted twice with phenol/choloform (1:1) and the aqueous layer was transferred to a sterile corex tube. RNA was precipitated by addition of 1/10th volume NaAc pH 5.2 and 2.5 volumes RNAase free ethanol, and incubated at -80°C for 30 minutes or overnight at -20°C. After spinning at 10 K for 30 minutes at 4°C, the RNA pellet was resuspended in 450 µl of sterile water, and transferred to an eppendorf tube for re-precipitation as above. The RNA was pelletted for 15 minutes, resuspended in an appropriate volume of water, and the concentration was determined by spectrophotometery. RNA samples were stored at -20°C.

2.3.17 Isolation of RNA from Tissue Samples

RNA from tissue samples was isolated using the acid guanidium thiocyanate method (Chomcvynski and Sacchi, 1987).

2.3.18 Isolation of $(A)_n^+$ RNA

Oligo-d(T) cellulose beads were prepared by soaking in 0.1 M NaOH and rinsed thoroughly in MQ H_20 before resuspending in pre-wash solution (0.5 M NaCl, 0.4 M Tris-HCl pH 7.5, 10 mM EDTA, 0.5% SDS). 500 µg cytoplasmic RNA was precipitated and resuspended in 100 µl MQ H_20 before denaturation at 80°C for 2 minutes and snap cooling on ice. 100 µl resuspension solution (1 M NaCl, 0.4 M Tris-HCl pH 7.5, 20 mM EDTA, 1% SDS) was added to the RNA before transferring to 100 µl oligo-dT cellulose beads. After incubation at room temperature for 5 minutes the beads were washed twice with 0.5 M NaCl / 10 mM EDTA and the $(A)_n^+$ RNA eluted twice in 150 µl MQ H_20. RNA was precipitated by addition of 650 µl ethanol and 8 µl 4M NaCl, resuspended in MQ H_20, and stored at -20°C.

2.3.19 RNA Gels for Northern Blot Analysis

1.3% agarose gels for northern blot analysis were prepared by dissolving 3.9 g agarose in 225 ml MQ water. Once the gel solution had cooled to 60°C, 25 ml 10 X MOPS and 15 ml 20% formaldehyde (freshly prepared by dissolving 4 g paraformaldehyde in 20 ml MQ H_2O) were added before pouring the gel.

$(A)_n$ RNA samples (approximately 10 µg) were prepared for electrophoresis as follows:

11.25 µl	RNA + MQ H_2O
5 µl	10 X MOPS
8.75 µl	formaldehyde (37%, pH 4.5)
25 µl	deionised formamide

RNA samples were denatured by heating at 65°C for 15 minutes and snap cooled on ice. GLB was before loading into wells. 5 µg *Eco* RI / *Hind* III lambda DNA markers were also loaded. Northern gels were run at 6V/cm gel length in 1 X MOPS until the bromophenol blue dye had reached the bottom of the gel.

2.3.20 Transfer of Northern Blot

The lane containing the lambda DNA markers was removed from the gel stained in ethidium bromide for 45 minutes and destained in water overnight before photographing under medium wavelength UV light. The remainder of the gel was blotted onto Nytran nylon membrane using capillary transfer. Two pieces of whatman 3MM paper were pre-wetted in 20 X SSC and placed over a platform so that the edges of the paper were submerged in 20 X SSC. The gel was placed wells facing down on the damp Whatman paper (avoiding bubbles) and parafilm placed around the gel to avoid short circuiting. The nylon membrane was pre-wetted in 20 X SSC and then carefully placed over the gel such that no air bubbles were trapped between the filter and the gel. Two pieces of Whatman paper, pre-wetted in 20 X SSC, were then placed on top of the membrane, followed by a 2 cm stack of dry Whatmann paper. A 5 cm stack of paper towels was then added and a glass plate placed on top of the paper towels. A 0.8 kg weight was placed on the glass plate and transferred for at least 24 hours. Following transfer, the RNA was covalently cross-linked to the filter by exposure to 120 mJ of UV radiation in a Stratagene UV Stratalinker^{TM} before pre-hybridization.

2.3.21 Ribonuclease Protection

RNAase protection analysis were essentially performed using the method Krieg and Melton (1987). Transcription reactions contained 240 μCi alpha-^{32}PdUTP and 100,000 counts/minute of single stranded probe was added to each RNA sample. RNAase digestion products were separated on a 7 M urea / 5% TBE polyacrylamide gel and visualised using autoradiography and phosphorimager analysis.

Hesx1a and Hesx1b specific riboprobes for RNAase protection were generated from the pHESa and pHESb plasmids respectively. pHESa was prepared by ligating a 928 bp *Xmn* I/*Sca* I restriction fragment (*Sca* I site at position 628) derived from the Hesx1a pBluescript clone (4.2.1) into *Sma* I linearized pT7T319U vector. pHESb plasmid was constructed from a 926 bp *Sca* I/*Pst* I fragment (*Pst* I site at position 770), derived from the Hesx1b pBluescript clone (4.2.1), ligated into *Sma* I/*Pst* I linearized pT7T319U. The *Xmn* I and *Sca* I restriction sites in the Hesx1a and Hesx1b pBluescript clones respectively, were

due to the presence of artifactual 5' sequences generated during library construction (4.2.1). The Hesx1a (643 bp) and Hesx1b (761 bp) riboprobes were transcribed from *Esp* I (position 26, Fig 4.2A) linearized templates using T7 RNA polymerase.

The murine gluteraldehyde phosphate dehydrogenase (mGAP) loading control probe was generated by SP6 RNA polymerase transcription of a 300 bp cDNA clone linearized with *Bam* H1 (Rathjen *et al.*, 1990).

2.3.22 Synthesis of Radioactive DNA Probes

Single stranded DNA probes were prepared using a Gigaprime labelling kit with 10-60 µCi alpha-^{32}PdATP. To remove unincorperated label, probe reactions were diluted to 100 µl with MQ H_2O, loaded onto a Sephadex G-50 column and centrifuged at 3.2 K for 4 minutes. The mGAP loading control for northern blot analysis was prepared by oligolabelling of the 300 bp cDNA clone plasmid (Rathjen *et al.*, 1990).

2.3.23 Synthesis of Radioactive RNA Probes

Riboprobes were synthesized as described by Krieg and Melton (1987). RNAase protection probes contained 240 µCi alpha-^{32}PUTP and 0.15 mM UTP. Northern probes contained 60 µCi alpha-^{32}PUTP and 12 µM UTP. Unincorperated label was removed using a Sephadex G-50 column as for DNA probes (2.3.22). Antisense *Oct-4* riboprobes for northern blots were prepared by T3 RNA polymerase transcription of *Hind* III linearized template (see 2.1.7).

2.3.24 Hybridization of Radioactive Probes to Nylon and Nitrocellulose Filters

Hybridization reactions were carried out in a Hybaid Hybridization Oven.

(i) <u>DNA probes</u>

Prehybridization solution contained 1 M NaCl, 40% deionized formamide, 1% SDS, 10 % PEG, 50 mM Tris-HCl pH 7.4 and 5 X Denhardt's. Filters were prehyridized for a minimum of 4 hours at 42^{o}C in 10 ml prehybridization solution/filter. DNA probes (2.3.22)

were boiled for 2 minutes with 2.5 mg sonicated salmon sperm DNA, snap-cooled on ice before adding to cylinders. Filters were probed overnight at 42°C.

(ii) RNA probes

Northern filters were prehybridized for a minimum of 4 hours at 65°C in 10 ml hybridization solution containing 5 X SSC, 60% formamide, 5 X Denhardt's, 20 mM sodium phosphate pH 6.8, 1% SDS, 100 µg/ml sonicated salmon sperm DNA and 100 µg/ml denatured tRNA. RNA probes (2.3.23) were denatured by heating at 85°C for 2 minutes, snap-cooled on ice and added to cylinders. Hybridization reactions were carried out overnight at 65°C and washed in 2 X SSC / 1% SDS at 50°C for 3 X 15 minutes and then in 0.2 X SSC / 1% SDS at 75°C for 1 hour.

2.3.25 Synthesis of Digoxygenin (DIG) Labelled RNA Probes

DIG labelled riboprobes were prepared run-off transcription. The antisense *Oct-4* riboprobes for *in situ* hybridization was prepared by T3 RNA polymerase transcription of *Hind* III linearized template (see 2.1.7). Reactions contained 0.5 µg linearized plasmid, 40 mM Tris-HCl pH 7.5, 6 mM $MgCl_2$, 2 mM spermidine-HCl, 10 mM DTT, 1 unit RNAasin, 1 X DIG labelling mix and 20 units RNA polymerase in a total volume of 20 µl. Transcription reactions were incubated at 37°C for 2 hours. Riboprobes were precipitated by addition of 80 µl MQ H_20, 20 µl 100 mM EDTA, 10 µl 3 M Sodium Acetate pH 5.2, 250 µl 100% ethanol and 20 µg glycogen and incubation overnight at -20°C. After centrifugation for 15 minutes, the probe was resuspended in 100 µl of RNAase free MQ H_2O with 0.5 units RNAasin and stored at -20°C.

2.3.26 Dot Blot Assay for DIG Probe Reactivity

Nitrocellulose membrane was soaked in 20 X SSC and air-dryed. A 2 µl sample of DIG labelled probe (2.3.26) was serially diluted 10 fold to 1/10000, heated at 100°C for 5 min and snap-cooled on ice. 1 µl of each dilution was spotted onto the filter and bound by UV cross-linking. The filter was washed for 1 minute in buffer 1, 30 minutes in buffer 2 (blocking reagent), and 1 minute in buffer 1. Anti-DIG Fab antibody-alkaline phosphatase

conjugate was diluted to 1:5000 in buffer 1 and incubated with filter for 30 minutes. Unbound antibody was removed by washing 2 x 15 minutes in buffer 1 and the membrane equilibrated in buffer 3 for 2 minutes. Alkaline phosphatase reaction solution was prepared by mixing 4.5 µl NBT and 3.5µl BCIP with 1 ml buffer 3. This solution was added to filter and developed for 30 minutes in the dark. The reaction was terminated by washing the membrane in buffer 4 for 5 minutes. Probes used for *in situ* hybridization were reactive to a 1/1000 dilution.

2.3.27 *In Situ* Hybridization of Cell Monolayers

Cell monolayers were rinsed twice with PBS and fixed overnight in 4% paraformaldehyde (PFA) / PBS solution. The following day, the cells were rinsed twice on ice with PBT (0.1% Tween 20 / PBS solution) and dehydrated on ice by the addition of 25% methanol / PBS for 5 minutes, 50% methanol / PBS for 5 minutes, 75% methanol / PBS for 5 minutes and stored in 100% methanol for up to 3 months. Cells were rehydrated to PBS using the methanol on ice series, and rinsed three times at room temperature in PBT. Cells were then washed for 3 x 20 minutes with RIPA buffer and fixed in 4% PFA / 0.2% gluteraldehyde in PBT for 20 minutes. Plates were rinsed 3 X 5 minutes in PBT and washed for approximately 10 minutes in 1:1 hybridization buffer:PBT (hybridization buffer, 50% deionized formamide, 5 X SSC, 0.1% Tween-20, 50 µg/ml heparin). Cells were washed for a further 10 minutes in hybridization buffer, and prehybridized at 68°C in a sealed container containing 50% formamide in hybridization buffer containing 10 µg/ml denatured salmon sperm DNA and 10 µg/ml denatured yeast tRNA for 1-6 hours. DIG-labelled riboprobes were denatured at 80°C for 10 minutes, added to fresh hybridization buffer containing denatured salmon sperm DNA and yeast tRNA, and allowed to hybridize overnight at 68°C. Probes were typically diluted 1:150 for the hybridization step. The following day, cell monolayers were washed in 50% formamide / 2 X SSC / 0.1% Tween 20 for 5 minutes and 3 X 30 minutes at 68°C. The cells were then washed three times at room temperature in TBST (136 mM NaCl, 2.7 mM KCl, 25 mM Tris-HCl pH 7.5, 0.1% Tween 20) followed by incubation in 10% FCS / TBST for one hour at room temperature. Anti-digoxygenin Fab

fragments-AP conjugate (Boehringer Mannheim) was diluted 1:2000 and 1:250 respectively in 1% FCS / TBST and incubated with cells overnight at 4°C. Next day, cells were washed for 3 X 5 minutes and 3 X 45 minutes in TBST and 3 X 10 minutes in AP buffer. The cells were developed in AP buffer containing 0.45 mg/ml NBT and 0.18 mg/ml BCIP in the dark until purple staining appeared (3-48 hours). The staining reaction was terminated by rinsing several times with PBT / 1 mM EDTA.

2.3.28 Containment Facilities

All manipulations involving viable organisms which contained recombinant DNA were carried out in accordance with the regulations and approval of the Australian Academy of Science Committee on Recombinant DNA and the University Council of the University of Adelaide.

2.3.29 Phosphorimager Analysis and Autoradiograph Scanning

Gels and filters were exposed to Storage Phosphor Screens (Molecular Dynamics) and processed using a Molecular Dynamics PhosphorImager running the ImageQuant software package. Autoradiographs were scanned using a Hewlett Packard *ScanJet IIcx* scanner running the DeskScan II 2.0 software package. PhosphorImager and DeskScan files were manipulated using the AdobePhotoshop™ and Mac Draw Pro programs, and printed using a Hewlett Packard LaserJet4 printer.

2.4 Tissue Culture Methods

2.4.1 Maintenance of ES cells

ES cells were maintained on gelatinised 10 cm petri dishes (Corning or Falcon) in complete ES medium at 37°C in 10% CO_2. Cells were harvested as follows: cells were washed twice in PBS, incubated with trypsin (1 ml) at 37°C and transferred into 4 ml complete ES cell medium. The cells were spun at 1.2 K for 5 minutes, gently resuspended in 10 ml complete ES medium and re-seeded at clonal density (30-50 fold dilution). ES cells were re-seeded every 3-4 days.

2.4.2 Differentiation of ES Cells

(i) Preparation of X cells

ES cells were seeded at clonal density in X cell medium (with or without LIF). X cell colonies were detected within 24-48 hours after induction. X cells were maintained in X cell medium, and passaged and harvested as described for ES cells (2.4.1).

(ii) Spontaneous induction

ES cells were seeded at 1000 cells/cm^2 in incomplete ES medium. After the initial 72 hours of culture, the medium was changed daily, and the cells were cultured for a further 5-6 days and harvested (2.4.1).

(iii) Retinoic Acid Induction

ES cells were seeded at 1 X 10^4 cells/cm^2 in incomplete ES medium containing 10 U/ml recombinant LIF (Amrad) and allowed to adhere overnight. The following day the medium was replaced with identical medium containing either 2 μM (low) or 10 μM (high) retinoic acid, and the cells cultured for a further 72 hours with daily changes of (RA containing) medium. The cells were cultured for a further 2-4 days in incomplete ES cell medium (without LIF) before harvesting (2.4.1).

(iv) 3-methoxybenzamide Induction

ES cells were seeded at 1000 cells/cm^2 in incomplete ES medium containing 100 units/ml recombinant LIF (Amrad) and allowed to adhere overnight. The following day, the

medium was replaced with identical medium containing 5 mM 3-methoxybenzamide (MBA), and the cells induced for 48 hours. The cells were cultured for a further 2-4 days before harvesting (2.4.1).

(v) DMSO Induction

ES cells were seeded at 1000 cells/cm^2 in incomplete ES medium containing 100 units/ml recombinant LIF (Amrad) and allowed to adhere overnight. On the following day the culture medium was replaced with identical medium plus 1.5% DMSO. DMSO induction was maintained for 72 hours with daily medium changes. Following induction, cells were cultured in medium without LIF for a further 3-4 days and harvested (2.4.1).

2.4.3 Preparation of Embryoid Bodies

Embryoid bodies were generated in bacterial dishes by culture of 1 X 10^6 ES cells in incomplete ES cell medium.

2.4.4 Preparation of Stable ES Cell Lines

MBL-5 ES cells (Pease *et al*, 1990) were transfected using a modification of the calcium phosphate co-precipitation method of Chen and Okayama (1987). 2 x 10^6 ES cells (passage 14) were seeded into gelatinised 10 cm petri dishes and cultured in complete ES medium for 24 hours. 20 µg vector DNA was mixed with 4 µg pPGKNeo (Whyatt *et al.*, 1993) before addition to 500 µl 0.25 M $CaCl_2$. 500 µl 2 X BBS was added dropwise to this solution, and the solution vortexed and incubated at room temperature for 20 minutes. A mock transfection control (without DNA) was also carried out. The ES cells were washed with 1:1 Hams F12 medium:complete ES medium and re-fed with this medium mixture. The DNA solution was added dropwise to the cell medium, with gentle mixing. The cells were transfected for 24 hours at 37°C under 3% CO_2, before replacing medium with ES cell selection medium. Selection medium was replaced every 48 hours until no visible mock transfected cells remained (about 10 days). Resistant colonies were picked by rinsing cells with PBS and local application of trypsin, and the cells were pipetted several times before seeding successively into single wells of 24- and 6- well trays and finally into 10 cm petri-dishes. For long term storage, the cells were washed twice in PBS, harvested, washed in 5

ml complete ES cell medium, and resuspended in 800 µl 90% FCS / 10% DMSO. The cells were then transferred to a 1ml NUNC tube, and stored at -80°C. Frozen ES cell lines were recultured by thawing the cells at 37°C before transferring into 10 ml complete ES medium. The cells were pelletted by centrifugation at 1 K for 3 minutes, resuspended in 10 ml complete ES medium, and seeded (typically at a 1/5 to 1/20 dilution).

CHAPTER 3: HOMEOBOX GENE EXPRESSION IN EMBRYONIC STEM CELLS AND DIFFERENTIATED DERIVATIVES

CHAPTER 3: HOMEOBOX GENE EXPRESSION IN EMBRYONIC STEM CELLS AND DIFFERENTIATED DERIVATIVES

3.1 Introduction

Embryonic stem (ES) cells provide an unique model system for the analysis of pluripotent stem cell biology. ES cells can be propagated in the undifferentiated state as monocultures, by culture in the presence of the cytokine DIA/LIF (Smith *et al.*, 1988; Williams *et al.*, 1988), or induced to differentiate along alternative pathways by withdrawal of DIA/LIF or by exposure to chemical inducers (Smith, 1991). ES cells reintroduced into the blastocyst are able to colonize all of the differentiated tissues which constitute the developing embryo, including the germline (Robertson, 1987; Bradley *et al.*, 1984). This demonstrates that ES cells cultured *in vitro* retain their full biological potential and therefore provide an accessible source for the isolation of genes which regulate developmental decisions during early murine ontogeny.

Extensive analyses of homeobox gene expression and function during murine embryogenesis have demonstrated that homeobox gene products are key components in the establishment of pattern formation and regional identity during development (Hunt and Krumlauf, 1992, McGinnis and Krumlauf, 1992). However, these studies have focussed almost exclusively on post-implantation stages of development, and relatively little is known about the developmental role or expression of homeobox genes during early embryogenesis. Homeobox genes expressed by pluripotent cells are of particular interest as they are likely to control developmental decisions in the early embryo, and provide molecular markers for the identification of distinct pluripotent cell types in the early embryo (Chapter 1).

This chapter describes the isolation of thirteen homeobox sequences (including three novel sequences) from ES cell cDNA using PCR. The expression of these homeobox genes in undifferentiated ES cells, and a range of differentiated ES cells, was investigated by northern blot analysis and *in situ* hybridization on cell monolayers. The results presented in this chapter indicate that homeobox genes are likely to be expressed during early

embryogenesis, and may provide molecular markers for the identification of discrete sub-populations of pluripotent cells.

3.2 PCR Screen of Homeobox Gene Expression in ES Cells

3.2.1 PCR Screen using *Antennapedia* and *Engrailed* Primers

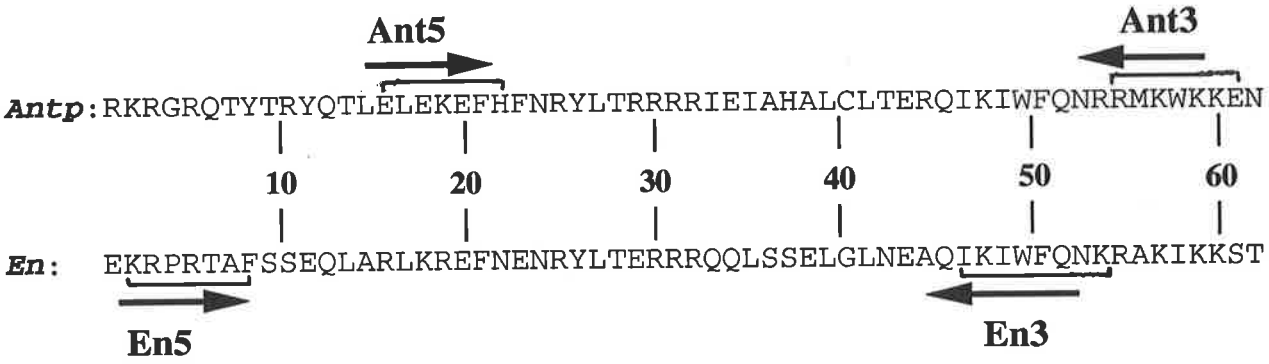
cDNA from MBL-5 ES cells (Pease *et al.*, 1990) maintained in the presence of DIA/LIF was screened for homeobox gene expression using PCR. An initial screen was carried out with degenerate oligonucleotide primers, designed to amplify homeobox genes belonging to the *Antennapedia* (Ant5, Ant3) and *Engrailed* (En5, En3) sequence classes (Fig. 3.1). These primers were based on conserved sequence motifs in the *Drosophila Antennapedia* and *Engrailed* homeodomain sequences (Kamb *et al.*, 1989), and were only able to amplify a subset of murine *Antennapedia* and *Engrailed* class genes due to base pair mismatches at the 3' end of the primers. PCR products were separated on a 8% polyacrylamide gel, endfilled, subcloned into *Sma* I linearized pT7T319U vector and sequenced in both orientations using the T3 and T7 primers.

Nine clones derived from the Ant5 and Ant3 primers were analysed. Seven of these clones contained the *Hox-a7* (Colberg-Poley *et al.*, 1985b) partial homeobox sequence, and single clones of the *Hox-a1* (Baron *et al.*, 1987) and *Hox-b1* (Frohman *et al.*, 1990; Murphy and Hill, 1991) genes were isolated. Fifteen clones generated from the *Engrailed* primers were sequenced. Eleven of these clones encoded *En-2* (Joyner and Martin, 1987), a murine homolog of the *Drosophila Engrailed* gene, and three clones contained the partial homeobox sequence of the *msx-2* gene (Holland, 1991; Monaghan *et al.*, 1991). The remaining clone encoded a novel homeodomain sequence which contained eight of the nine highly conserved amino acid residues which span this region of the homeodomain (Fig. 3.2A). This novel homeobox gene was named *Hesx1* (***Hesx1***-**H**omeobox gene expressed in **ES** cell; assigned by the International Committee on Mouse Gene Nomenclature).

Figure 3.1 Design of the *Antennapedia* and *Engrailed* PCR primers.

A. The *Antennapedia* (*Antp*) and *Engrailed* (*En*) homeodomain sequences which correspond to the Ant5, Ant3, En5 and En3 degenerate PCR primers are shown. **B.** The nucleotide sequences for the Ant5, Ant3, En5 and En3 primers. N = A, C, G or T. / indicates positions of degeneracy.

A



B

***Eco* RI E L E K E F H**

Ant5: GAATTC GAA/G C/TTC GAA/G AAA/G GAA/G TTC/T CAT

***Hind* III E K K W K M R**

Ant3: AAGCTT C/TTC C/TTT C/TTT A/GTG C/TTT CAT A/G/TCG

K R P R T A F

En5: AAA/G A/CGN CCA A/CGN ACN GCA/C TT

K N Q F W I K I

En3: TT A/GTT C/TTG A/GAA CCA GAT C/TTT A/G/TAT

3.2.2 PCR Screening using Class Specific Degenerate Homeobox Primers

A more comprehensive PCR screen for homeobox genes expressed by ES cells was carried out using the 5'mANT and 5'mEN upstream degenerate primers, which were directed against the ELEKEF (*Antennapedia, Deformed, labial, Abdominal B, Hox 1.5* (*Hox-a3*), *Hox 2.4* (*Hox-b8*) classes) and the RPRTAF (*Engrailed* class) sequence motifs which are conserved in homeobox genes belonging to these classes (Scott *et al.*, 1989; Fig. 3.3). These 5' primers were used in conjunction with a single degenerate 3' primer which was complimentary to the WFQNRR motif (3'ALL) found in most homeodomain sequences (Scott *et al.*, 1989; Fig 3.3). *Eco* RI sites were included at the 5'end of the primers to facilitate cloning. PCR products were digested with *Eco* RI, isolated from a 8% polyacrylamide gel, ligated into pT7T319U and sequenced from both strands.

Twenty-one clones amplified by the 5'mANT and 3'ALL primers were sequenced. Four homeobox gene sequences which were not isolated in the previous screen were identified: *msx-1* (four clones; Hill *et al.*, 1989), *Hex* (five clones; Bedford *et al.*, 1993), *MMoxA* (four clones; Murtha *et al.*, 1991) and *MMoxB* (eight clones; Murtha *et al.*, 1991). *MMoxA* and *MMoxB* are highly related, and encode identical homeodomain sequences which are discriminated by 13 nucleotide differences spanning the known homeobox region.

Twenty-two clones generated by the 5'mEN and 3'ALL primers were analysed. The homeobox sequences isolated were *MMoxB* (nine clones), *msx-1* (three clones), *Hesx1* (two clones), *En-2* (one clone), *msx-2* (one clone) and *Gsh-1* (Singh, *et al.*, one clone). Two novel homeobox sequences, termed *Hesx2* (four clones) and *Hesx4* (one clone) were also identified using these primers (Fig. 3.2B). These homeodomain sequences encode 9 of the 11 highly conserved amino acid residues spanning this region of the homeodomain (as defined by Scott *et al.*, 1989; Fig. 3.2B), confirming their identity as homeodomain sequences.

Figure 3.2 Sequence comparison of the *Hesx1*, *Hesx2* and *Hesx4* partial homeodomain sequences.

A. The *Hesx1* partial homeodomain sequence is compared to the homeodomain consensus sequence (Scott *et al.* 1989), and the XANF-1 (Zariasky *et al.*, 1992) and Antennapedia (*Antp*; McGinnis *et al.*, 1984; Scott and Weiner, 1984) homeodomains. *Hesx1* residues which match the homeodomain consensus are shown in bold. The dashes in the XANF-1 and *Antennapedia* sequences indicate positions of identity with *Hesx1*. **B.** Comparison of the *Hesx2* and *Hesx4* partial homeodomain sequences with the DMHOMEO, *nkch-4* (Jagla *et al.*, 1993) and *Antennapedia* homeodomains. Residues which match the consensus homeodomain sequence are shown in bold.

A

		HOMOLOGY
CONSEN:	---Q---L---F----Y-----R---A--L-L---Q	
***Hesx1*:**	TQNQVEVLENVFRVNCYPGIDIREDLAQKLNLEEDR	
XANF-1:	-RS-I-I--------S-----V--E--S--A-D---	80%
***Antp*:**	-RY-TLE--KE-HF-R-LTRRR-IEI-HA-C-T-RQ	40%

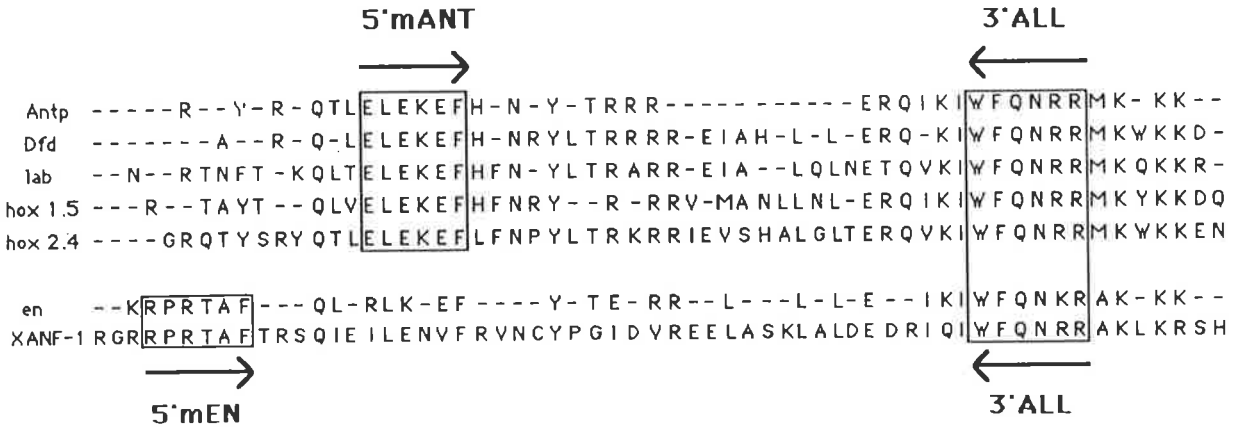
B

		HOMOLOGY	
CONSEN:	---Q---L---F----Y-----R---A--L-L---Q-KI	***Hesx2***	***Hesx4***
***Hesx2*:**	TNHQIYELEKRFLYQKYLSPADRDQIAQQLGLTNAQVIT	100%	59%
***Hesx4*:**	TAQQVLELERRFVFQKYLAPSERDGLAARLGLANAQVVT	59%	100%
DMHOMEO:	TNHQIFELEKRFLYQKYLSPADRDEIAASLGLSNAQVIT	87%	62%
***nkch-4*:**	TNQQIFELEKRFLYQKYLSPADRDEIAGGLGLSNAQVIT	85%	62%
***Antp*:**	TRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKI	41%	44%

Figure 3.3 Design of the class specific degenerate homeobox PCR primers.

A. The consensus sequences (as defined by Scott *et al.*, 1989) of the *Antennapedia* (Antp), *Deformed* (Dfd), *Labial* (lab), *Hox-a3* (hox 1.5), *Hox-b8* (hox 2.4) and *Engrailed* (en) homeodomain classes are shown. The XANF-1 homeodomain sequence (Zariasky *et al.*, 1992) is also included. Dashes indicate the positions of amino acid residues which are not conserved. The conserved sequences which correspond to the unique 5' primers (5'mANT and 5'mEN) and the common 3' primer (3'ALL) are boxed. **B.** The nucleotide sequences for the 5'mANT, 5'mEN and 3'ALL primers. N = A, C, G or T. / indicates positions of degeneracy.

A



B

***Eco* RI E L E K E F**

5'mANT: TA GAATTC GAA/G C/TTN GAA/G AAA/G GAA/G TT

***Eco* RI R P R T A F**

5'mEN: TA GAATTC A/CGN CCN A/CGN ACN GCN TT

***Eco* RI R R N Q F W**

3'ALL: TA GAATT CCG/T NCG/T A/GTT C/TTG A/GAA CCA

3.2.3 Sequence Class Allocation of Partial Homeodomain Sequences Isolated in the PCR Screen

Comparison between homeodomain sequences allows grouping of homeobox genes into sequence classes (1.5.4). As a general rule, homeobox genes within the same class share at least 60% identity across the entire homeodomain sequence (Kappen *et al.*, 1993), and are less than 50% homologous to members of other homeodomain classes (Scott *et al.*, 1989)

Of the thirteen homeobox sequences which were isolated from ES cells, the following seven genes belonged to established sequence classes: *En-2* (*Engrailed* class), *Hox-a1* and *Hox-b1* (*labial* class; Scott, *et al.*, 1989; Frohman *et al.*, 1990), *Hox-a7* and *Gsh-1* (*Antennapedia* class; Scott, *et al.*, 1989; Burglin, 1994), *msx-1* and *msx-2* (*msh* class; Holland, 1991).

To determine the relationship between *Hesx1*, *Hesx2* and *Hesx4* and existing homeodomain sequences, the partial homeodomain sequences were compared to the GenBank, EMBL, PIR and Swiss-Prot data bases. The *Hesx1* partial homeodomain shared 72% homology with the *Xenopus* homeobox gene XANF-1 (Zariasky *et al.*, 1992), and had less than 50% identity with other homeodomain sequences (Fig. 3.2A). This indicated that *Hesx1* and XANF-1 define a novel class of homeodomain sequences. The *Hesx2* partial homeodomain sequence shared 85% homology with the *Drosophila* homeobox gene *nkch-4* (Jagla *et al.*, 1993) and 87% homology with an unpublished *Drosophila* homeodomain sequence (GenBank Accn. No. DMHOMEO; Fig. 3.2B). Less than 62% identity between *Hesx2* and other homeodomain sequences was observed, suggesting that *Hesx2*, together with the related *Drosophila* sequences, establish a new class of homeodomain sequences. The *Hesx4* partial homeodomain sequence shared greatest similarity with the *nkch-4* and DMHOMEO sequences (62%; Fig. 3.2B). However, given that *Hesx4* was less than 60% related to the *Hesx2*, this degree of identity is probably not sufficient to include *Hesx4* within the *Hesx2 / nkch-4 /* DMHOMEO homeodomain sequence class. Therefore, the *Hesx4* homeodomain may define a novel sequence class. It must be noted, however, that comparison of the entire *Hesx1*, *Hesx2* and *Hesx4*

homeodomain sequence is required to confirm their status as members of novel homeodomain sequence classes.

Hex, *MMoxA* and *MMoxB* also encode relatively diverged homeodomain sequences. Previous analysis of the entire *Hex* homeodomain sequence (Bedford *et al.*, 1993) identified only 55% homology with its closest relative *Hlx* (Allen *et al.*, 1991), indicating that *Hex* defined a novel sequence class. Similarly, *MMoxA* and *MMoxB* shared 100% identity (across the PCR amplified region) with the human homologs GBX1 (Matsui *et al.*, 1993) and GBX2 (Matsui *et al.*, 1994), and the chicken homeobox gene *Ovx1* (Fainsod and Gruenbaum, 1989), but were not significantly related to other homeodomain sequences. Therefore, of the thirteen homeobox genes isolated from ES cells, six genes (including three novel sequences) encoded relatively diverged homeodomain sequences which were not significantly related to existing homeodomain sequences.

3.3 Developmental Expression of Homeobox Genes Isolated in the PCR Screens

None of the homeobox genes isolated in the PCR screen have been shown to be expressed in the pluripotent cells of the mouse embryo, or in embryonic stem (ES) cells. However, several of the genes which were isolated are expressed in temporally and spatially restricted patterns at later stages of development. *Hox-a1* and *Hox-b1* are expressed in distinct patterns in the central nervous system and in similar distinct regions of the pre-somitic mesoderm (Murphy and Hill, 1991). *Hox-a7* has restricted expression in the neural tube, spinal ganglia and sclerotomes (Mahon *et al.*, 1988). *En-2* is expressed in the anterior neural folds at 8.0 days post-coitum (p.c.) (Davis and Joyner, 1988), and in the developing midbrain and the midbrain-hindbrain junction at 12.5 days p.c. (Davidson *et al.*, 1988). *msx-2* and *msx-1*, are expressed in related patterns during histogenesis of the eyes, teeth, heart and early limb buds (Holland, 1992; Hill *et al.*, 1989). Although *Hex* expression in the embryo has not been mapped, expression has been detected in a range of myeloid and lymphoid progenitor cell lines (Bedford *et al.*, 1993). The embryonic expression of the remaining homeobox genes which were isolated in the PCR screen is not known.

3.4 Analysis of Homeobox Gene Expression in ES cells and Differentiated ES Cell Derivatives

The expression of homeobox genes isolated using the PCR screen was investigated in undifferentiated ES cells, and in a range of differentiated ES cell derivatives. This expression analysis was undertaken for two reasons. Firstly, characterization of gene expression in undifferentiated and differentiated ES cells *in vitro* should provide some insight into the expression pattern in pluripotent cells during early embryogenesis. Secondly, as a low background level of differentiated cells is unavoidably present in ES cell cultures grown in the presence of DIA/LIF, northern blot analysis is required to determine whether the homeobox genes isolated in the PCR screen were expressed by undifferentiated or differentiated ES cells.

3.4.1 Differentiation of ES Cells

Differentiated ES cells used in this analysis were induced using three different regimes:

(i) Spontaneous differentiation of ES cells

Culture of ES cells in the absence of DIA/LIF results in the spontaneous differentiation of ES cells into a range of terminally differentiated cells of ill-defined phenotype (Smith, 1991; Fig. 3.4A). Spontaneously differentiated ES cell cultures also contained nests of residual pluripotent cells which can be identified morphologically and by expression of *Oct-4* (see 3.4.2). The perseverance of these pluripotent cells is thought to result from feedback expression of DIA/LIF by the surrounding differentiated cells (Rathjen *et al.*, 1990).

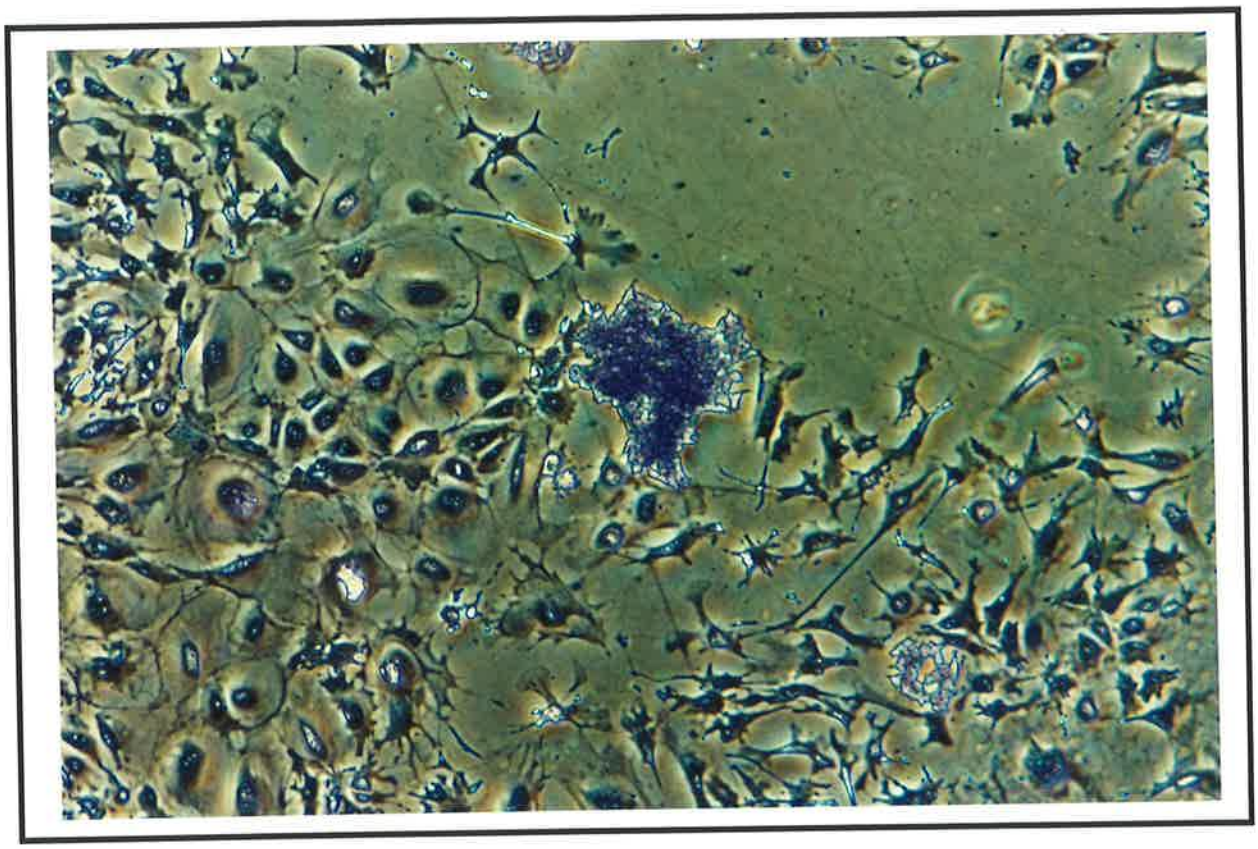
(ii) Chemical induction of ES cell differentiation

Chemical induction of ES cell differentiation was carried out using retinoic acid (RA), 3-methoxybenzamide (MBA) or dimethylsulfoxide (DMSO). In my hands, retinoic acid (RA) induction was the most efficient method of ES cell differentiation and generated two

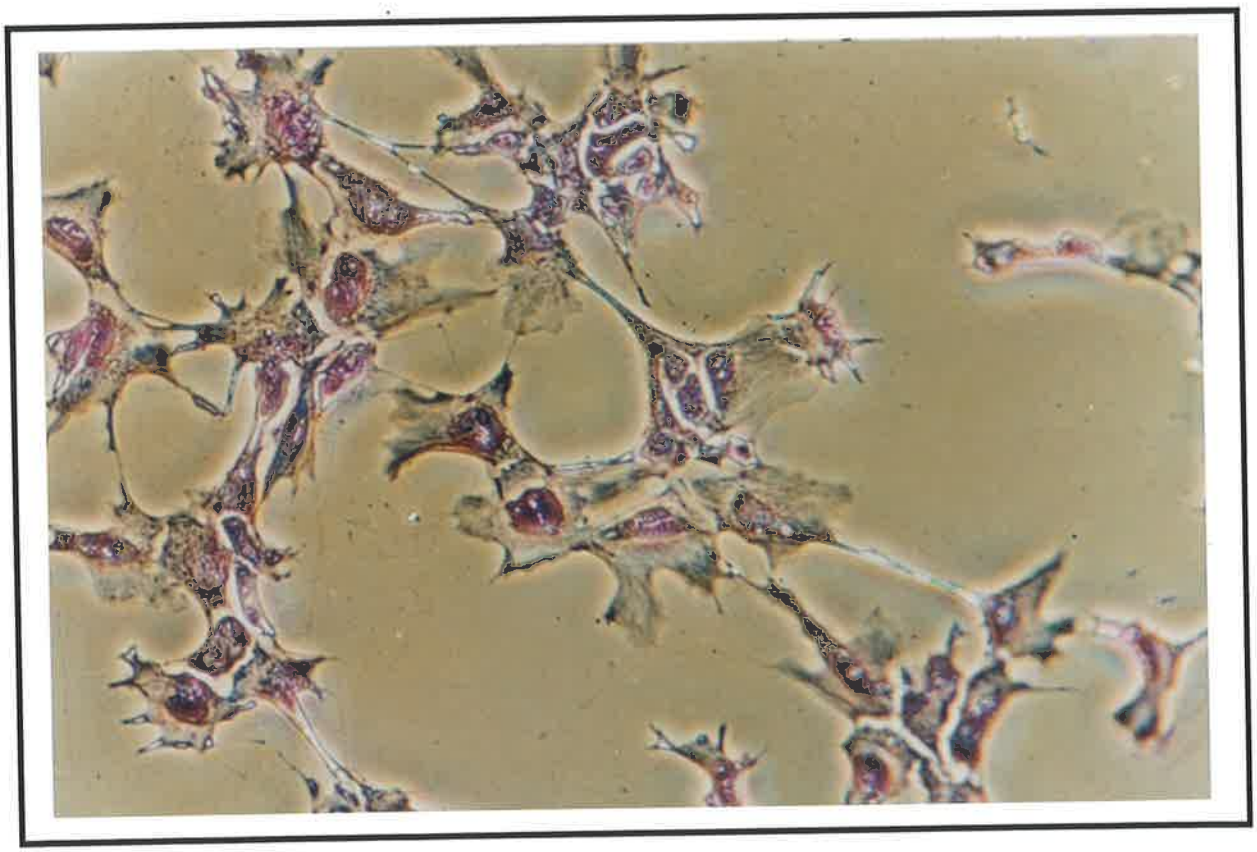
Figure 3.4 Spontaneous and chemical differentiation of ES cells.

Differentiated cell types generated by ES cell culture in the absence of LIF (A) or by exposure to retinoic acid (B), 3-methoxybenzamide (C) and dimethylsulfoxide (D). Cells were cultured for 5 days, stained with Giemsa and photographed under phase contrast optics at 20 X magnification.

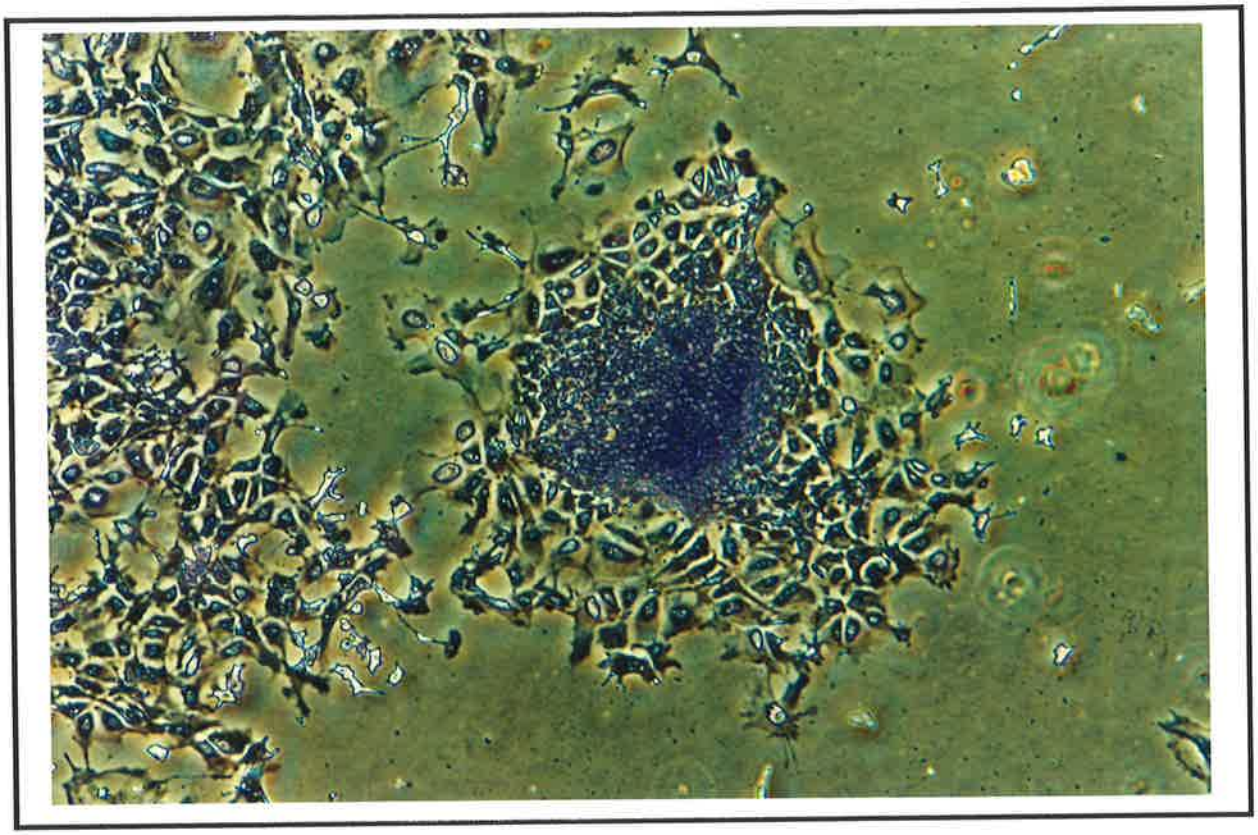
A



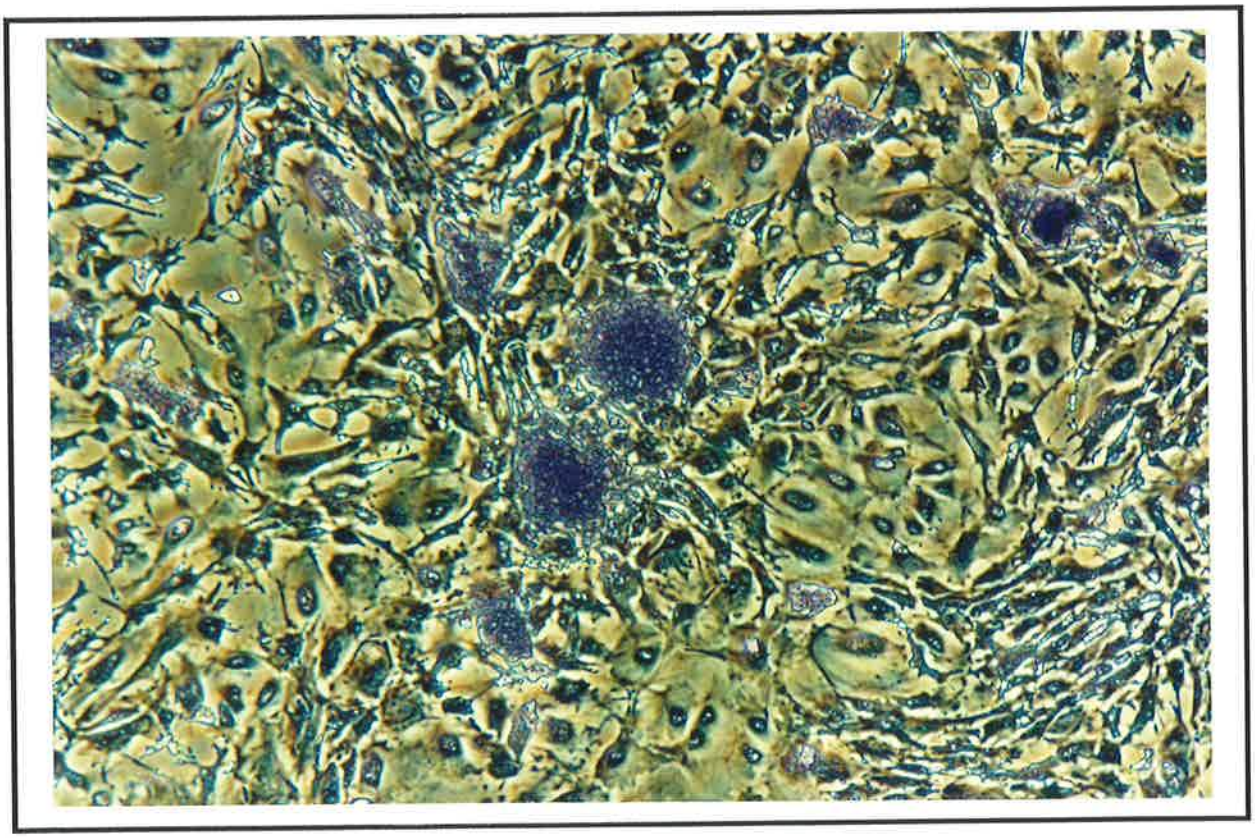
B



C



D



different terminally differentiated cell types; large fibroblast cells and variable amounts of refractile parietal yolk sac-like cells (Fig. 3.4B; Smith, 1991). DMSO- and MBA-induced differentiation were less efficient and result in the formation of terminally differentiated cells with nests of residual pluripotent cells (Fig. 3.4C,D). As in spontaneous differentiation, the maintenance of pluripotent cells in these cultures is thought to result from feedback expression of LIF by the differentiated cells (Rathjen *et al.*, 1990).

(iii) Formation of X cells in MED2 conditioned media

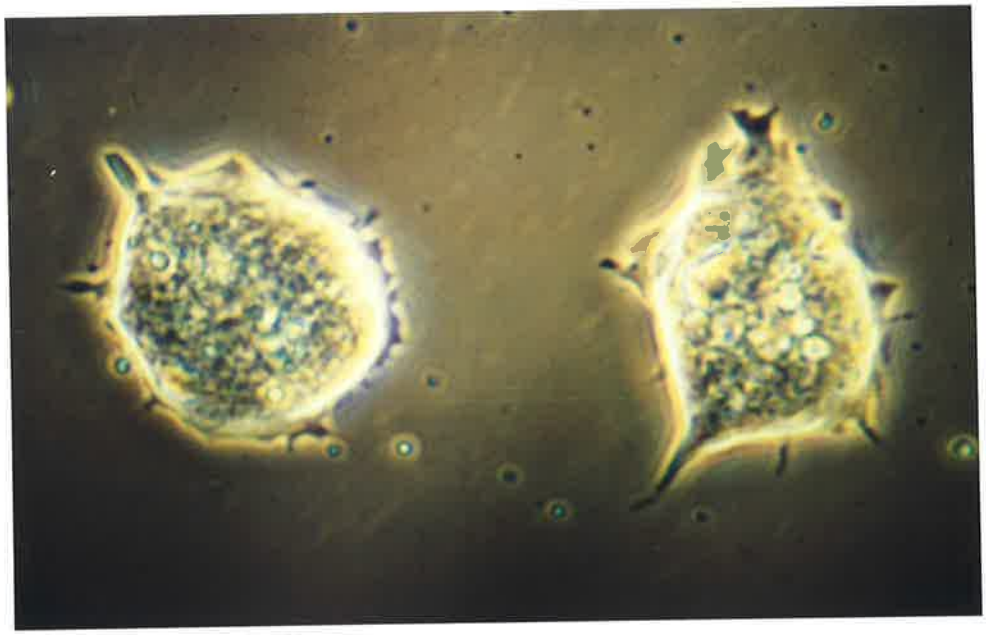
Culture of ES cells in media conditioned by a transformed human hepatic cell line (MED2) induces differentiation of ES cells into an homogeneous monolayer of flat cells, termed X cells (Joy Rathjen , unpublished results; Bettess, 1993; Fig 3.5). The conversion of ES cells to X cells can be carried out either in the presence or absence of DIA/LIF. ES cells seeded in the absence of DIA/LIF (and in the presence of conditioned media) and passaged once before harvesting were termed X-1 cells (- = -DIA/LIF, 1 = 1 passage). Induction of X cells in the presence of DIA/LIF and subsequent culture for one or three passages were termed X+1 and X+3 cells respectively. The X cells generated using the three methods were morphologically indistinguishable.

Several lines of evidence, based on analysis of gene expression and cellular differentiation, have suggested that ES cells and X cells are closely related but distinct pluripotent cell types (Bettess, 1993). Firstly, culture of X cells in the absence of conditioned media and in the presence of DIA/LIF induces the reversion of X cells to ES cells (Fig. 3.5). Secondly, ES cells and X cells express the pluripotent cell markers *Oct-4* (1.8), alkaline phosphatase (Adamson, 1988) and SSEA-1 (Solter and Knowles, 1978). However, comparison of gene expression in ES cells and X cells using PCR differential message display (Liang and Pardee, 1992) has detected subtle molecular differences between ES cells and X cells (Tom Schulz, personal communication). In addition, analysis of marker gene expression in ES cells and X cells has indicated that X+3 cells may represent an *in vitro* equivalent of the primitive ectoderm, and that X+1 cells are an intermediate cell type in the ES to X cell conversion (Bettess, 1993). Thirdly, X cells, like

Figure 3.5 Conversion of ES cells to X cells.

Culture of ES cells in the presence of MED2 conditioned media, either in the presence or absence of LIF, induces the formation a homogeneous layer of flattened cells termed X cells. X cells can be reverted to ES cells by culture in the absence of MED2 and the presence of LIF. The ES and X cells were cultured for 3 days and photographed under phase contrast optics at 50 X magnification.

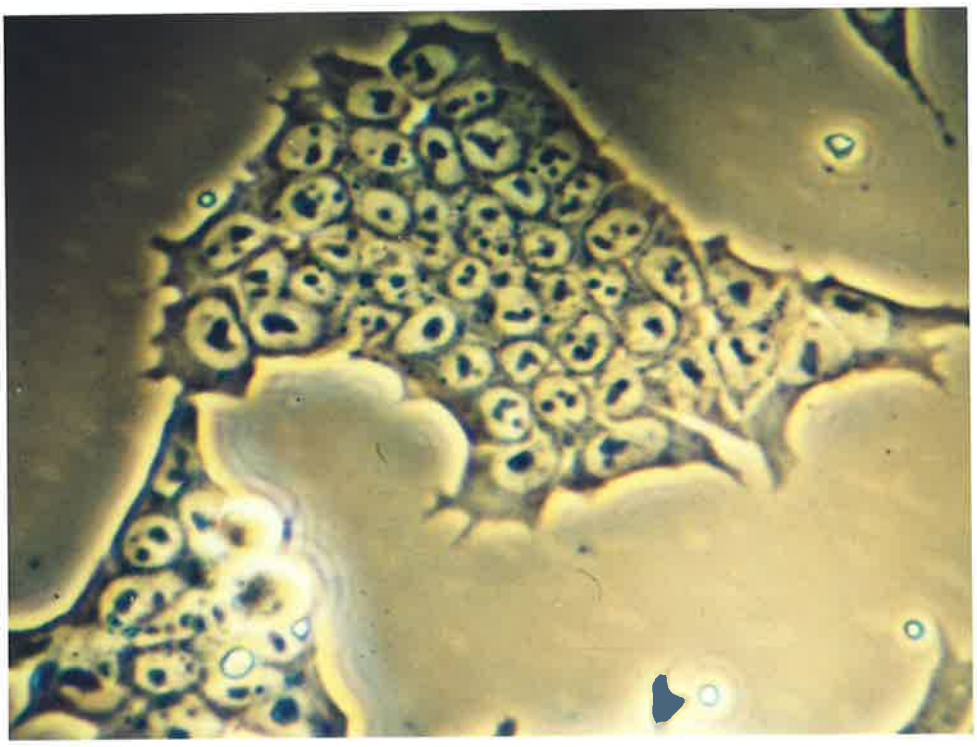
ES Cells



- MED2
+ LIF

+ MED2
+/- LIF

X Cells



ES cells, are able to form cellular aggregates, termed embryoid bodies (Robertson, 1987) which are capable of generating complex differentiated tissues such as cardiac muscle, neurons and blood (Doetschmann *et al.*, 1985). However, in comparison with ES cell embryoid bodies, X cell embryoid bodies differentiate more readily into cardiac muscle, and have a reduced propensity to form neurons (Julie-Anne Lake, unpublished results). These lines of evidence indicate that the ES to X cell transition *in vitro* may reflect the formation of primitive ectoderm from the inner cell mass in the early murine embryo.

3.4.2 Identification of Residual Pluripotent Cells using *Oct-4*

The perseverance of pluripotent cells within the differentiated cell population can be assessed by expression of the homeobox gene *Oct-4* (Scholer *et al.*, 1990a; Rosner *et al.*, 1990; Yeom *et al.*, 1991; Okamoto *et al.*, 1990). During development, *Oct-4* expression is restricted to pluripotent cell lineages, which include the inner cell mass and primitive ectoderm (Rosner *et al.*, 1990; Yeom *et al.*, 1991; Scholer *et al.*, 1990b). ES cells, which are derived from the inner cell mass, express *Oct-4* at high levels (Scholer, 1991), and this expression is rapidly down-regulated during ES cell differentiation (Scholer *et al.*, 1990b; Okamoto *et al.*, 1990; Rosner *et al.*, 1990; Scholer *et al.*, 1990a). The relative levels of undifferentiated cells after spontaneous or chemical induction were determined by comparison of *Oct-4* expression levels in the uninduced and induced ES cell populations using northern blot and *in situ* hybridization analysis.

3.4.3 Northern Blot Analysis of Homeobox Gene Expression

Radioactive antisense riboprobes generated from linearized pT7T319U plasmid containing the PCR products, were used to probe northern blot filters carrying undifferentiated MBL-5 (Pease *et al.*, 1990) ES cell, X cell and induced ES cell RNA. This analysis allowed classification of the homeobox genes isolated in the PCR screen into six groups based on their expression pattern in undifferentiated and differentiated ES cells.

(i) Group I (*Hesx1*)

Two *Hesx1* transcripts of approximately 1.0 kb and 1.2 kb were detected in ES cells, X cells and the spontaneous-, MBA- and DMSO-induced cells (Fig. 3.6). *Hesx1* expression was not detected in ES cells induced to differentiate with RA. The absence of residual pluripotent cells resulting from RA induction, as implied by the lack of *Oct-4* expression, indicated that *Hesx1* expression was down-regulated during ES cell induction using RA. This expression pattern was consistent with a developmental role for *Hesx1* in pluripotent cells.

The detection of *Oct-4* expression in the spontaneous-, MBA-, and DMSO-induced cells indicated the presence of residual pluripotent cells in these cultures. Therefore, it is unclear whether the *Hesx1* expression detected in these cell populations was derived from the residual pluripotent cells, or from differentiated ES cells. However, the similar relative *Hesx1* and *Oct-4* expression levels in the spontaneous-, MBA-, and DMSO-induced cells, and the absence of *Hesx1* expression in the terminally differentiated RA induced cells, suggested that *Hesx1* expression may be restricted to the residual *Oct-4*$^{+}$ pluripotent cells. Therefore it seems likely that *Hesx1* expression was downregulated during ES cell differentiation induced by exposure to MBA or DMSO, or by withdrawal of LIF.

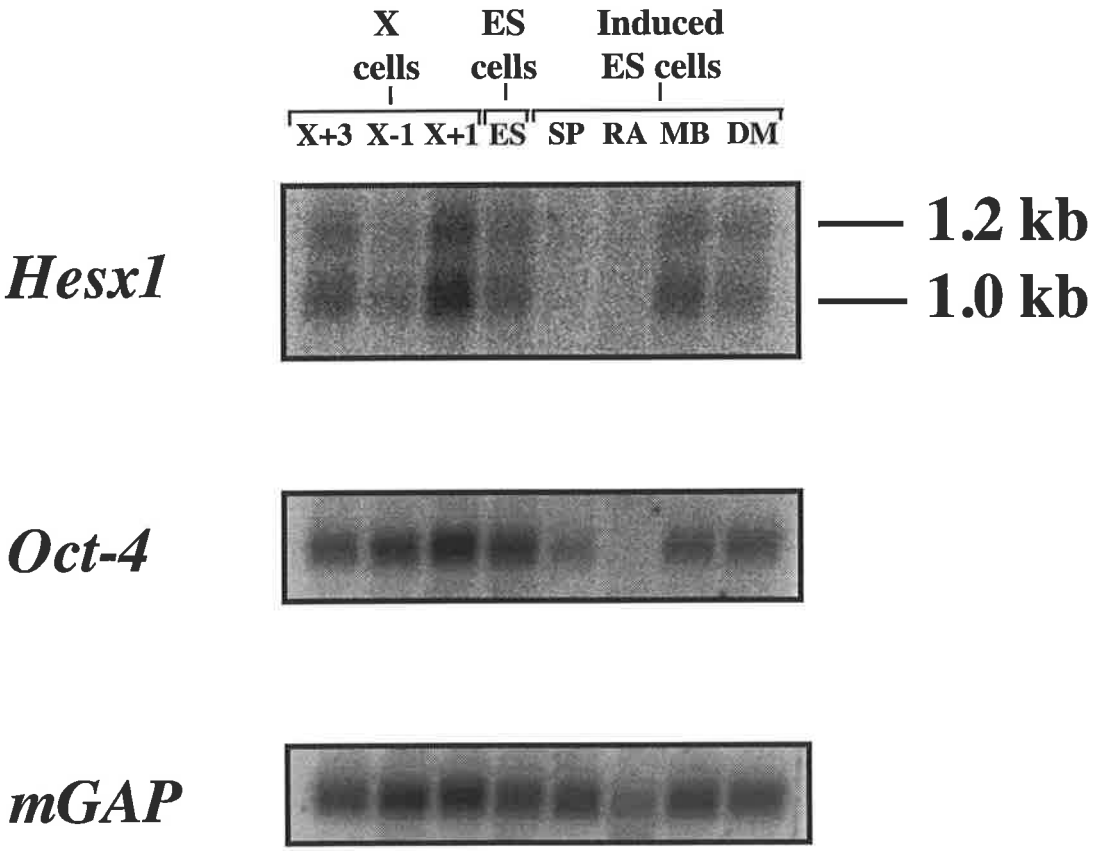
Hesx1 was expressed at different levels in the X+1, X-1 and X+3 cells. In comparison to uninduced ES cells, X+1 cells expressed an estimated 3-4 fold higher level of *Hesx1*. X+3 cells expressed *Hesx1* at about the same level as ES cells, and expression was downregulated in X-1 cells. More extensive northern blot analysis of X+1, X-1 and X+3 cells indicated that *Hesx1* expression in these cell cultures was variable (data not shown), although the reason for this variation is not known.

(ii) Group II (*Hesx2*)

Northern blot analysis of *Hesx2* expression is shown in Figure 3.7. *Hesx2* expression was not detected in uninduced ES cells, X cells or in spontaneous-, MBA- or DMSO-induced ES cells. A 2.1 kb *Hesx2* transcript were detected only in RA-induced ES cells. This pattern suggests that *Hesx2* expression is likely to be restricted to differentiated

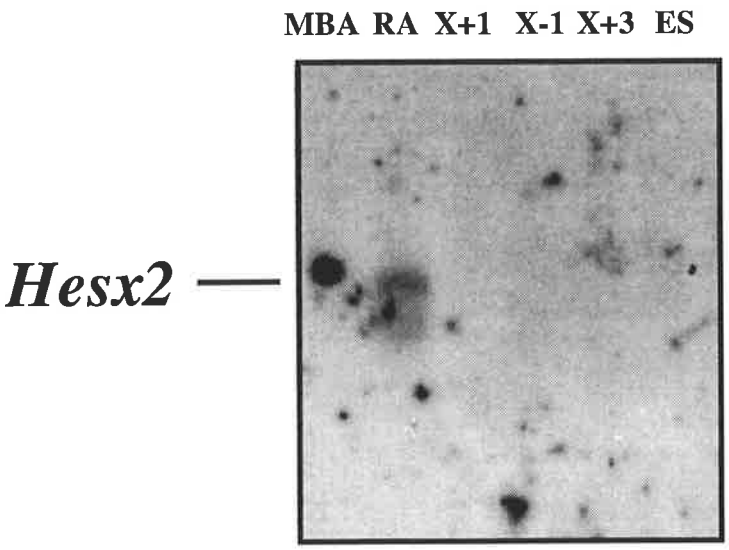
3.6 Northern blot analysis of *Hesx1* expression in uninduced and induced ES cells.

10 μg $(A)_n^+$ RNA isolated from ES cells, X cells and induced ES cells was electrophoresed on a 1.3% agarose gel and blotted. The northern filter was probed with an antisense riboprobe prepared from the *Hesx1* PCR product (3.2.1) and washed in 2 X SSC / 1% SDS at 50°C for 3 X 15 minutes followed by 0.2 X SSC / 1% SDS at 75°C for 1 hour. The northern blot was exposed in a phosphorimager cassette for 2 days. The filter was stripped and reprobed with *Oct-4* (2.3.23) and, subsequently, a *mGAP* loading control (2.3.22). X+3, X cells +LIF / 3 passages; X-1, X cells -LIF / 1 passage; X+1, X cells + LIF / 1 passage; ES, undifferentiated ES cells; SP, spontaneous (-LIF) induction; RA, retinoic acid induction; MB, 3-methoxybenzamide induction; DM, dimethylsulfoxide induction.


X cells
ES cells
Induced ES cells
X+3 X-1 X+1 ES SP RA MB DM
Hesx1
1.2 kb
1.0 kb
Oct-4
mGAP

3.7 Northern blot analysis of *Hesx2* expression in uninduced and induced ES cells.

10 μg $(A)_n^+$ RNA isolated from ES cells, X cells and induced ES cells was electrophoresed on a 1.3% agarose gel and blotted. The northern filter was probed with an antisense riboprobe prepared from the *Hesx2* PCR product (3.2.2) and washed in 2 X SSC / 1% SDS at 50°C for 3 X 15 minutes followed by 0.2 X SSC / 1% SDS at 75°C for 1 hour. The blot was exposed for 2 weeks at -80°C with an intensifying screen. X+3, X cells +LIF / 3 passages; X-1, X cells -LIF / 1 passage; X+1, X cells + LIF / 1 passage; ES, undifferentiated ES cells; RA, retinoic acid induction; MB, 3-methoxybenzamide induction.


MBA RA X+1 X-1 X+3 ES
Hesx2

tissues during embryogenesis, and is unlikely to play a developmental role in pluripotent cells. Given that *Hesx2* expression was not detected in ES cells, it seems likely that the isolation of *Hesx2* in the PCR screen was due to the low level of differentiated cells in the ES cell culture. Alternatively, *Hesx2* expression in ES cells may be below the level of detection using northern blot analysis.

(iii) Group III (*msx-1* and *msx-2*)

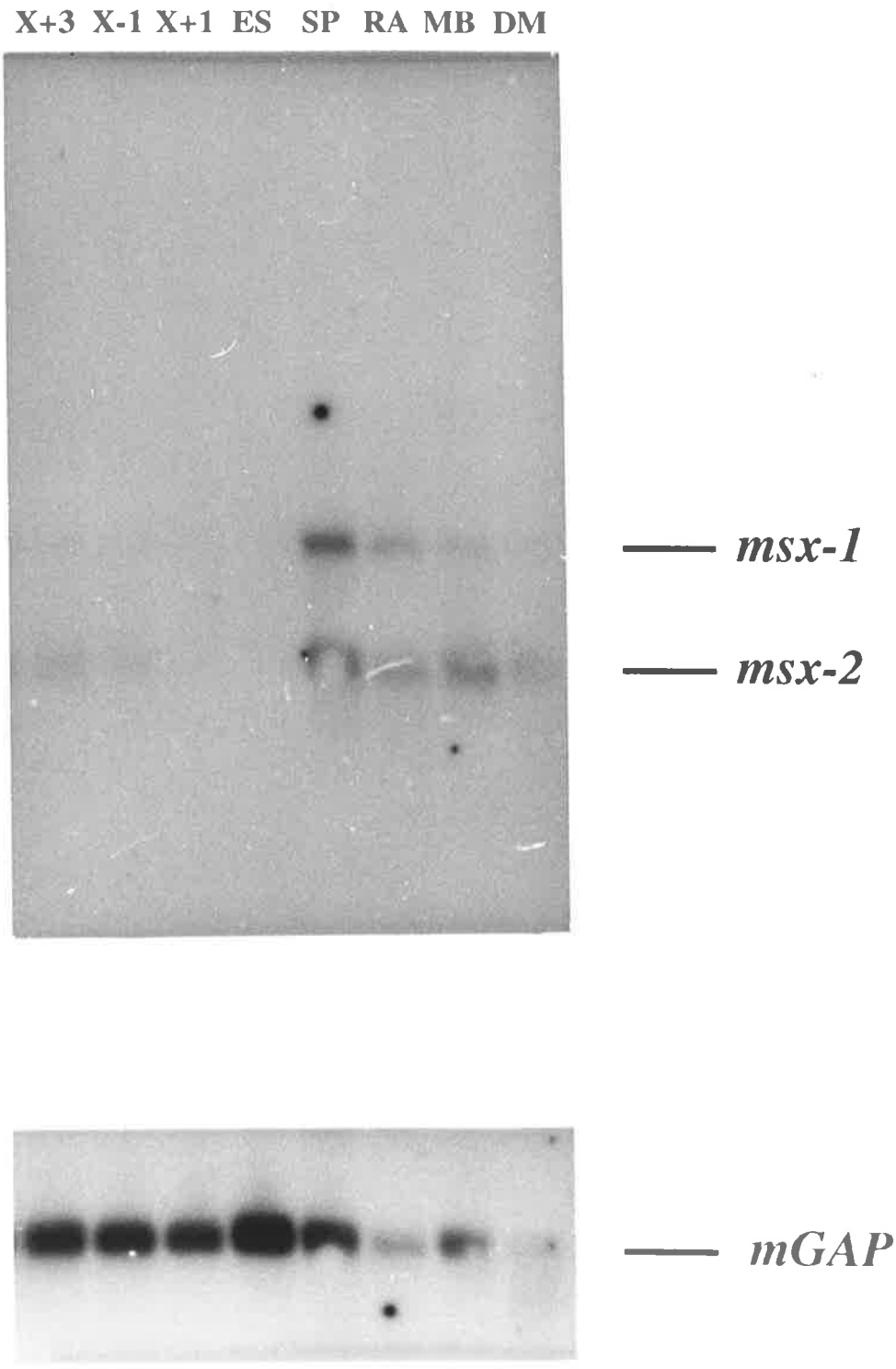
A northern blot carrying RNA derived from undifferentiated ES cells, X cells and chemically and spontaneously induced ES cells was probed with an antisense riboprobe generated from the *msx-2* PCR product amplified with the En5 and En3 PCR primers (3.2.1; Fig. 3.8). As the *msx-2* and *msx-1* homeobox sequences differ by only 18 nucleotides across the PCR amplified region, *msx-1* expression was also detected by cross-hybridization to the *msx-2* probe. The 3 kb *msx-1* transcript and 1.8 kb *msx-2* transcript were detected in spontaneously and chemically induced ES cells, and were not expressed at significant levels in undifferentiated ES cells or X cells. As the lanes corresponding to the chemically induced ES cells were underloaded (as determined by mGAP expression) the relative *msx-1* and *msx-2* expression in the RA-, DMSO- and MBA-induced cells was greater than is indicated by the intensity of the *msx-1* and *msx-2* bands. The induction of *msx-1* and *msx-2* expression in differentiated ES cells suggests that these genes are not expressed in the pluripotent cells of the early embryo. The restricted expression of *msx-1* and *msx-2* to differentiated cells may reflect the expression of these genes during later stages of development in the eyes, teeth, heart and early limb buds (Holland, 1992; Hill *et al.*, 1989).

(ii) Group IV (*Hesx4* and *Hex*)

Hex, *Hesx4* and *Oct-4* expression in uninduced and induced ES cells were compared using a common northern blot filter (Fig. 3.9). A 1.9 kb *Hex* transcript and a 2.5 kb *Hesx4* transcript were detected in uninduced ES cells, X cells and in spontaneous- and RA-induced ES cells. The expression of *Hex* and *Hesx4* in the spontaneously- and RA-induced ES cells

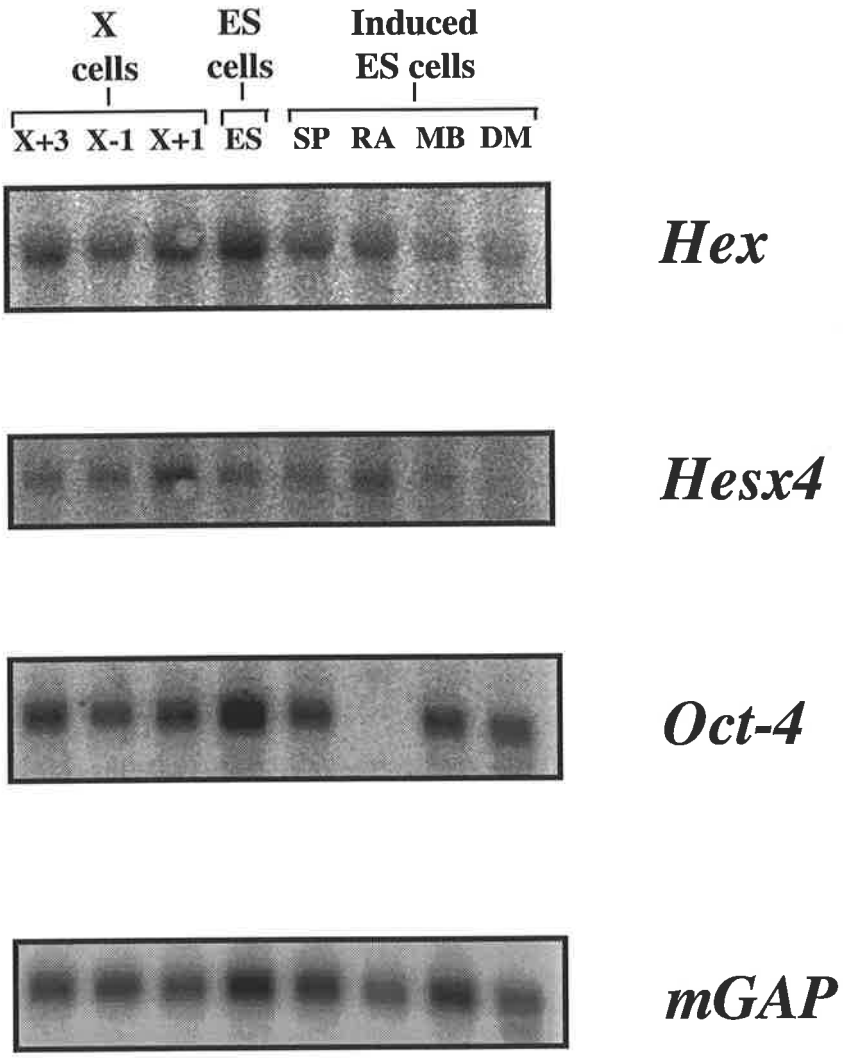
3.8 Northern blot analysis of *msx-1* and *msx-2* expression in uninduced and induced ES cells.

10 μg $(A)_n^+$ RNA isolated from ES cells, X cells and induced ES cells was electrophoresed on a 1.3% agarose gel and blotted. The northern filter was probed with an antisense riboprobe prepared from the *msx-2* PCR product (3.2.1) and washed in 2 X SSC / 1% SDS at 50°C for 3 X 15 minutes followed by 0.2 X SSC / 1% SDS at 75°C for 1 hour. The blot was exposed for 1 week at -80°C with an intensifying screen. Apart from the 1.8 kb *msx-2* signal, and additional 3 kb band corresponding to the *msx-1* transcript was detected due to cross-hybridization to the *msx-2* probe. The filter was stripped and reprobed with a mGAP loading control (2.3.22).


X+3 X-1 X+1 ES SP RA MB DM
msx-1
msx-2
mGAP

3.9 Northern blot analysis of *Hex* and *Hesx4* expression in uninduced and induced ES cells.

10 μg $(A)_n^+$ RNA isolated from ES cells, X cells and induced ES cells was electrophoresed on a 1.3% agarose gel and blotted. The northern filter was probed initially with a *Hex* antisense riboprobe, and subsequently with a *Hesx4* riboprobe. *Hex* and *Hesx4* riboprobes were generated from the *Hex* and *Hesx4* PCR products (3.2.2). After hybridization, the filter was washed in 2 X SSC / 1% SDS at 50°C for 3 X 15 minutes followed by 0.2 X SSC / 1% SDS at 75°C for 1 hour. The northern blot was exposed in a phosphorimager cassette for 2 days. The filter was stripped and reprobed with *Oct-4* (2.3.23) and, subsequently, a *mGAP* loading control (2.3.22). X+3, X cells +LIF / 3 passages; X-1, X cells -LIF / 1 passage; X+1, X cells + LIF / 1 passage; ES, undifferentiated ES cells; SP, spontaneous (-LIF) induction; RA, retinoic acid induction; MB, 3-methoxybenzamide induction; DM, dimethylsulfoxide induction.


X cells
ES cells
Induced ES cells
X+3 X-1 X+1 ES SP RA MB DM
Hex
Hesx4
Oct-4
mGAP

suggested that these genes are expressed by, and may play a developmental role in, differentiated cells during embryogenesis.

The MBA- and DMSO-induced cell populations, which contained a significant proportion of residual pluripotent cells (as indicated by *Oct-4* expression), expressed reduced levels of *Hex* and *Hesx4* compared with uninduced ES cells. Quantitative comparison between the relative *Hex* and *Oct-4* expression levels in ES cells and MBA- and DMSO-induced cells was carried out by volume integration. *Hesx4* was not included in this analysis due to the relatively high background on this filter. Comparison of the *Hex* versus *Oct-4* expression levels in MBA- and DMSO induced cells expressed as a percentage of the ES cell expression is shown in Figure 3.10. A greater relative decrease of *Hex* versus *Oct-4* expression was observed in the MBA- (13% versus 35%) and DMSO- (10% versus 28%) induced cells. It is not clear from this analysis whether the greater relative reduction in *Hex* signal compared with *Oct-4*, reflects *Hex* expression in a subset of the pluripotent cells which remained after MBA- and DMSO-induction, or a consistent reduction in *Hex* expression levels in all of the residual pluripotent cells. In either case, this expression pattern suggests that *Hex* and *Hesx4* may define sub-populations of pluripotent cells which arise during ES cell differentiation, and supports a developmental role for *Hex* and *Hesx4* during pluripotent cell differentiation in the early embryo.

Group V (*MmoxA/B*)

A single 2.3 kb transcript was detected by northern blot analysis of uninduced and induced ES cell RNA using a riboprobe derived from the *MMoxB* PCR fragment (Fig. 3.11). As there are only 13 nucleotide differences between *MMoxA* and *MMoxB* across the amplified region of the homeobox (Fig. 3.12), the *MMoxB* probe was expected to bind to both the *MMoxA* and *MMoxB* transcripts. This signal is therefore referred to as *MmoxA/B.*

MmoxA/B expression was detected at greatest levels in undifferentiated ES cells, and at slightly lower levels in the X+1 cells. Significantly reduced levels of *MmoxA/B* expression were detected in the X-1 and X+3 cells. The downregulation of *MmoxA/B* expression in the X-1 and X+3 cells, and the reduced expression in X+1 cells, indicated

Figure 3.10 Comparison of *Hex* and *Oct-4* expression in ES cells and MBA- and DMSO-induced ES cells.

Bands corresponding to *Hex* and *Oct-4* expression in undifferentiated ES cells, MBA- and DMSO-induced cells (Fig. 3.9) were quantitated by volume integration. Quantitation values for *Hex* and *Oct-4* expression MBA- and DMSO-induced cells are shown as a percentage of the ES cell expression value.

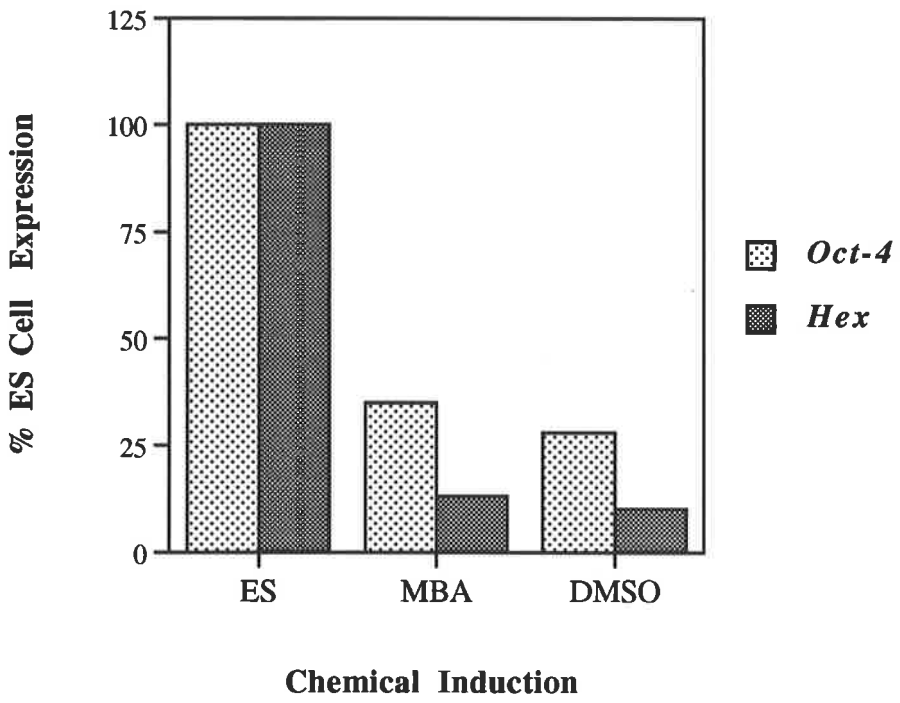

125
100
75
50
25
0
% ES Cell Expression
ES
MBA
DMSO
Chemical Induction
Oct-4
Hex

Figure 3.11 Northern blot analysis of *MmoxA/B* expression in uninduced and induced ES cells.

10 μg $(A)_n^+$ RNA isolated from ES cells, X cells and induced ES cells was electrophoresed on a 1.3% agarose gel and blotted. The northern filter was probed with an antisense riboprobe prepared from the *MMoxB* PCR product (3.2.2) and washed in 2 X SSC / 1% SDS at 50°C for 3 X 15 minutes followed by 0.2 X SSC / 1% SDS at 75°C for 1 hour. The northern blot was exposed in a phosphorimager cassette for 2 days. The filter was stripped and reprobed with *Oct-4* (2.3.23) and, subsequently, a *mGAP* loading control (2.3.22). X+3, X cells +LIF / 3 passages; X-1, X cells -LIF / 1 passage; X+1, X cells + LIF / 1 passage; ES, undifferentiated ES cells; SP, spontaneous (-LIF) induction; RA, retinoic acid induction; MB, 3-methoxybenzamide induction; DM, dimethylsulfoxide induction.

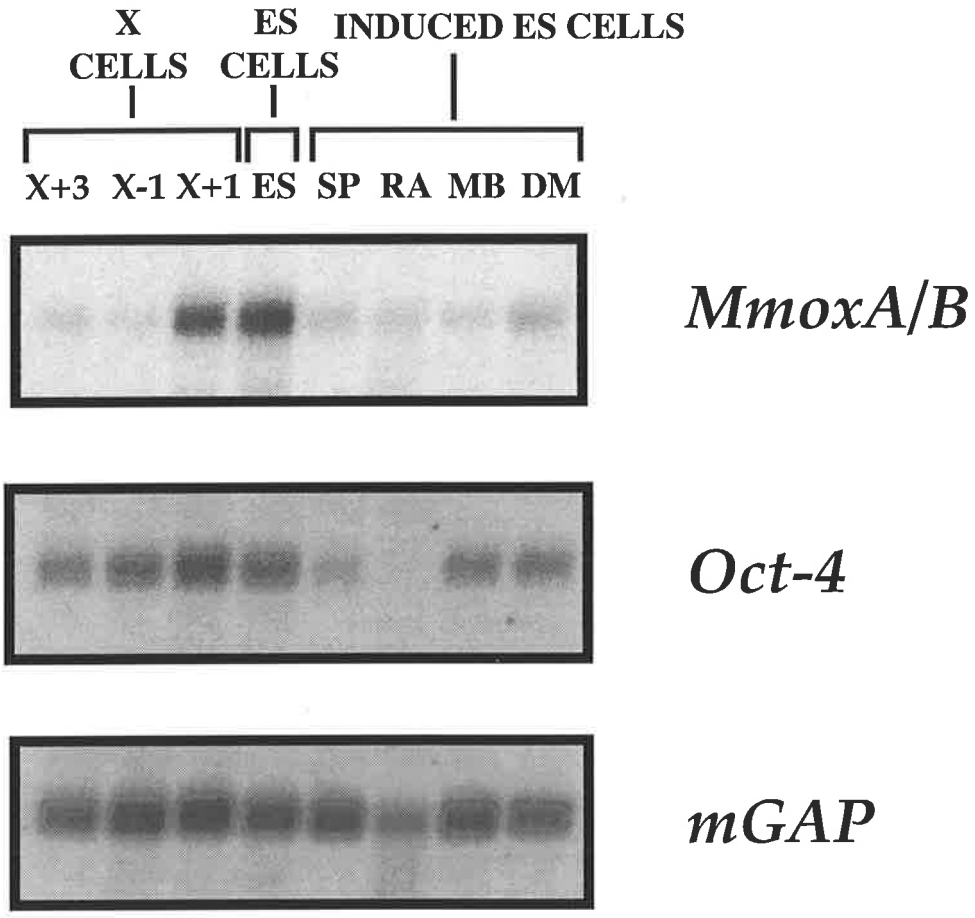

X CELLS
ES CELLS
INDUCED ES CELLS
X+3 X-1 X+1 ES SP RA MB DM
MmoxA/B
Oct-4
mGAP

Figure 3.12 Comparison of the *MMoxA* and *MMoxB* homeobox sequences.

Alignment of the *MMoxA* and *MMoxB* nucleotide sequences amplified by the 5'mANT and 3'ALL primers (3.2.2) spanning positions 61-141 of the homeobox. The positions of *MMoxB* residues were identical to *MMoxA* are indicated by dashes.

```
        61                                                  101
MmoxA:  CACTGCAAGAAATACCTGAGTCTGACAGAGCGCTCCCAGAT
MmoxB:  --------A--G-----CTCC-----C---------A-----

        102                                                 141
MmoxA:  CGCCCACGCCCTCAAGCTCAGTGAGGTGCAGGTGAAGATC
MmoxB:  ---------------A-----C---------A--A--A---
```

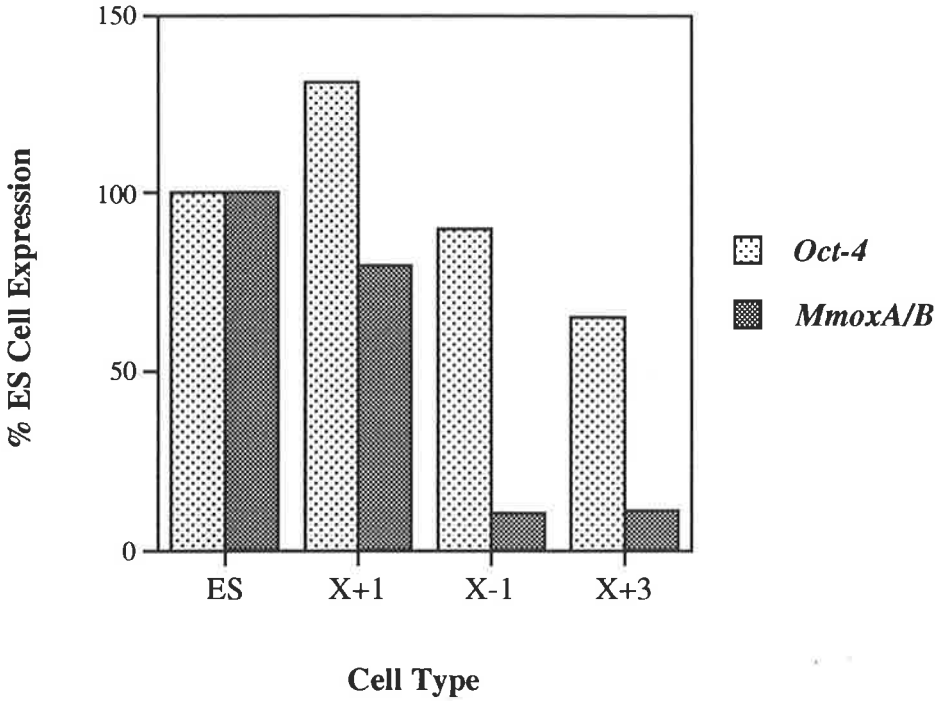
that transformation of ES cells to X cells resulted in a concomitant reduction in *MmoxA/B* expression. Quantitative comparison of *MmoxA/B* and *Oct-4* expression levels in ES cells and X cells was determined by volume integration (Fig. 3.13A). Relative *MmoxA/B* expression in ES cells, X+1, X-1 and X+3 cells was 100%, 54%, 7% and 14% respectively. In comparison, *Oct-4* expression in these cell types was 100%, 131%, 90% and 65%. This suggested the existence of two distinct pluripotent cell populations (as defined by *Oct-4* expression) which can be discriminated on the basis of *MmoxA/B* expression i.e. *Oct-4*$^+$/*MmoxA/B*$^+$ cells and *Oct-4*$^+$/*MmoxA/B*$^-$ cells.

The relative levels of *MmoxA/B* and *Oct-4* expression in uninduced ES cells and spontaneously and chemically induced ES cells were calculated by volume integration and expressed as a percentage of ES cell expression (Fig. 3.13B). The respective relative expression levels of *MmoxA/B* and *Oct-4* in the induced ES cells were as follows: spontaneous induction 15% versus 31%; RA induction 13% versus 6%; MBA induction 9% versus 59%; DMSO induction 22% versus 57%. The detection of *Oct-4* expression in the spontaneous-, MBA- and DMSO-induced cells was due the presence of residual undifferentiated ES cells, which could be identified by microscopic analysis of cell cultures. The low level of *MmoxA/B* (and *Oct-4*) expression in the RA induced ES cells indicated, firstly, that *MmoxA/B* expression was downregulated during ES cell differentiation with RA. Secondly, the lesser relative *MmoxA/B* expression compared with the *Oct-4* in the spontaneous-, MBA- and DMSO-induced cells suggested that only a proportion of the residual pluripotent cells (as defined by *Oct-4*) expressed *MmoxA/B*. This was consistent with the analysis of *MmoxA/B* expression in X cells which also suggested that *MmoxA/B* expression was expressed by a subset of (*Oct-4*$^+$) pluripotent cells. However, it is not clear whether the proposed *Oct-4*$^+$/*MmoxA/B*$^-$ cells generated during spontaneous and chemical ES cell induction, and the *Oct-4*$^+$/*MmoxA/B*$^-$ X-1 and X+3 cells, represent identical cell populations. Regardless of this distinction, the expression pattern of *MmoxA/B in vitro* suggested that *MMoxA* and/or *MMoxB* expression *in vivo* may be restricted to a subset of the pluripotent cells in the early mouse embryo, and is consistent with a developmental role for one or both of these genes during early development.

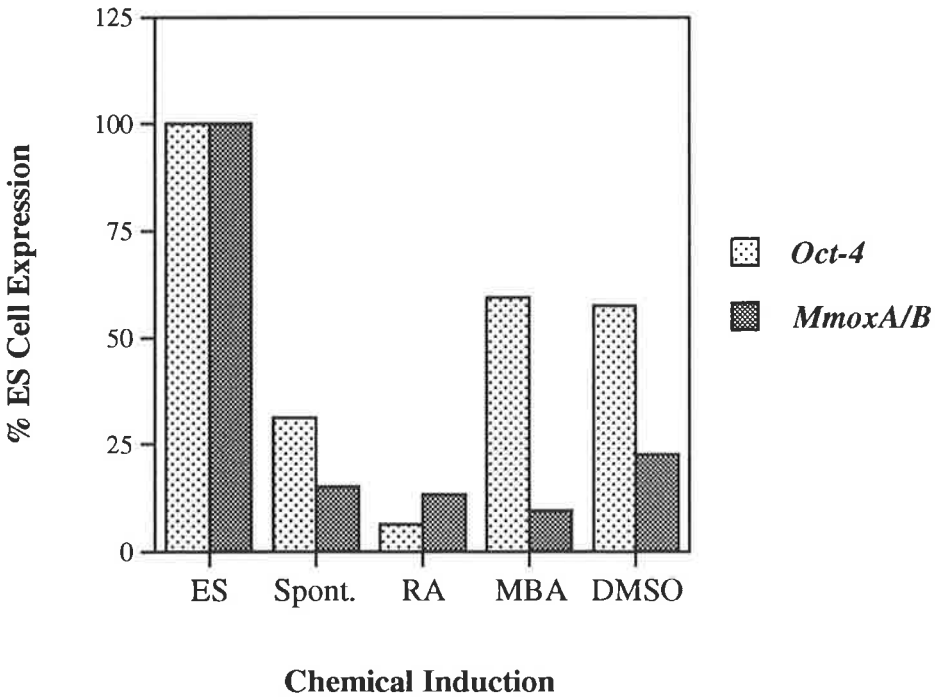
Figure 3.13 Comparison of *MmoxA/B* and *Oct-4* expression in ES cells and MBA- and DMSO-induced ES cells.

Bands corresponding to *MmoxA/B* and *Oct-4* expression in undifferentiated ES cells (A and B), X cells (A) and spontaneously and chemically induced ES cells (B) (Fig. 3.11) were quantitated by volume integration. Quantitation values for *MmoxA/B* and *Oct-4* expression are shown as a percentage of the ES cell expression value. X+3, X cells +LIF / 3 passages; X-1, X cells -LIF / 1 passage; X+1, X cells + LIF / 1 passage; ES, undifferentiated ES cells; SP, spontaneous (-LIF) induction; RA, retinoic acid induction; MB, 3-methoxybenzamide induction; DM, dimethylsulfoxide induction.

A


150
100
50
0
% ES Cell Expression
ES
X+1
X-1
X+3
Cell Type
Oct-4
MmoxA/B

B


125
100
75
50
25
0
% ES Cell Expression
ES
Spont.
RA
MBA
DMSO
Chemical Induction
Oct-4
MmoxA/B

(vi) Group VI (*Hox-a1*, *Hox-a7*, *Hox-b1*, *En-2* and *Gsh-1*)

Expression of *Hox-a1*, *Hox-a7*, *Hox-b1*, *En-2* and *Gsh-1* was unable to be detected in uninduced or induced ES cells by northern blot analysis. The isolation of these genes in the PCR screen may reflect their expression in ES cells at a level below the limit of detection using northern blot analysis. Therefore, further investigation into the expression of these genes in undifferentiated and differentiated ES cells requires the use of more sensitive techniques such as RNAase protection and quantitative PCR.

Although expression of *Hox-a1*, *Hox-a7*, *Hox-b1* and *En-2* have been detected in a range of tissues during middle and late embryogenesis (3.3), unlike *msx-1* and *msx-2*, expression of these genes was not detected in the spontaneously- and chemically-induced ES cells. This suggests that differentiated cells resembling these tissues were not generated using these differentiation regimes.

3.4.4 *In situ* Analysis of *MmoxA/B* and *Oct-4* Expression in Uninduced and Induced ES Cell Monolayers

The apparent restriction of *MmoxA/B* expression to a subset of the pluripotent cell population during ES cell differentiation was investigated further by *in situ* analysis of *MmoxA/B* and *Oct-4* expression in undifferentiated ES cells, and in ES cells induced to differentiate with MBA.

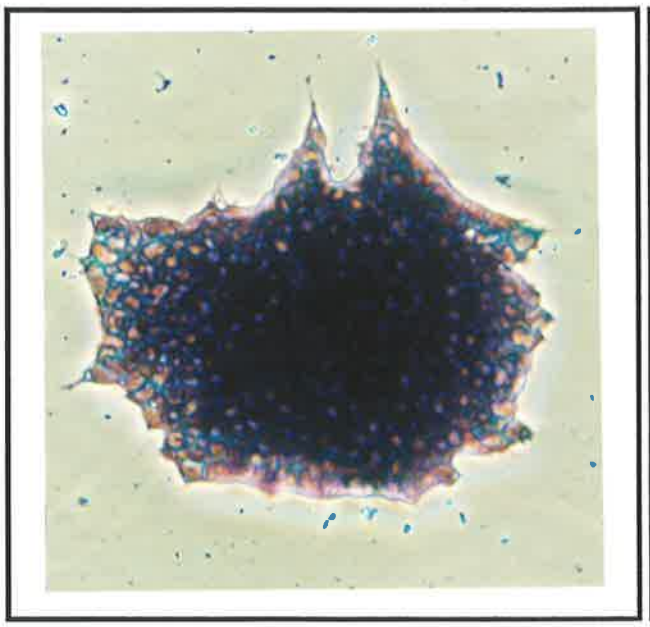
Undifferentiated ES cell colonies were probed with an antisense digoxygenin (DIG) labelled *Oct-4* riboprobe. This probe was identical to the *Oct-4* probe used for northern blot analysis of *Oct-4* expression. *Oct-4* expression (purple staining) was detected in all of the undifferentiated cells which constitute the ES cell colony (Fig. 3.14A,B). The specificity of the *Oct-4* signal was confirmed by *in situ* analysis using an *Oct-4* sense probe which did not result in staining (Fig. 3.14C,D).

MmoxA/B expression was detected using a DIG labelled antisense riboprobe generated from the *MMoxB* PCR clone. Like *Oct-4*, *MmoxA/B* expression was expressed by all of the undifferentiated cells within the ES cell colonies (Fig. 3.14E,F). *MmoxA/B* expression was not detected in the few differentiated cells which spontaneously arise during

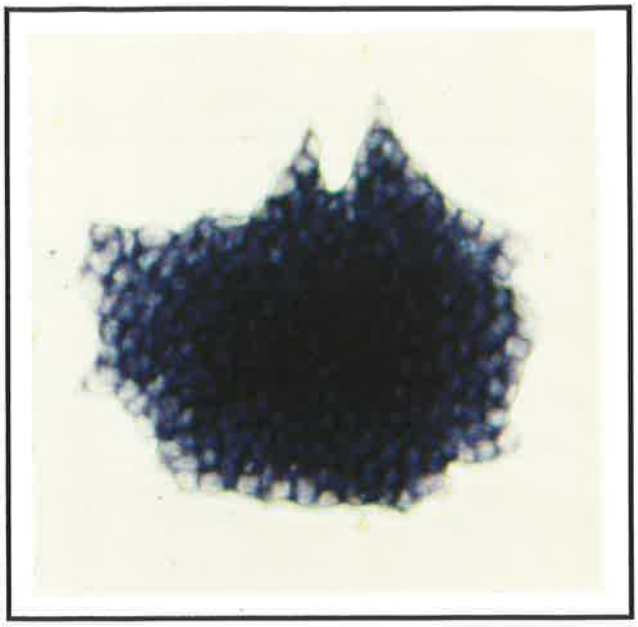
Figure 3.14 *In situ* analysis of *MmoxA/B* and *Oct-4* expression in undifferentiated ES cells.

Oct-4 expression (A, B) was detected using a 394 bp DIG labelled antisense riboprobe (2.3.25). *In situ* hybridization was also carried out using an *Oct-4* sense probe as a negative control. *MmoxA/B* expression (E, F) was detected using a DIG labelled antisense riboprobe derived from the *MMoxB* PCR product (3.2.2). ES cells were cultured for three days in the presence of LIF. The *Oct-4* and *MmoxA/B* plates were developed for 16 hours and 24 hours, respectively. Cells which express *Oct-4* or *MmoxA/B* are indicated by the purple staining. Photographs were taken under phase contrast (A, C, E) and brightfield optics (B, D, F) at 50 X magnification.

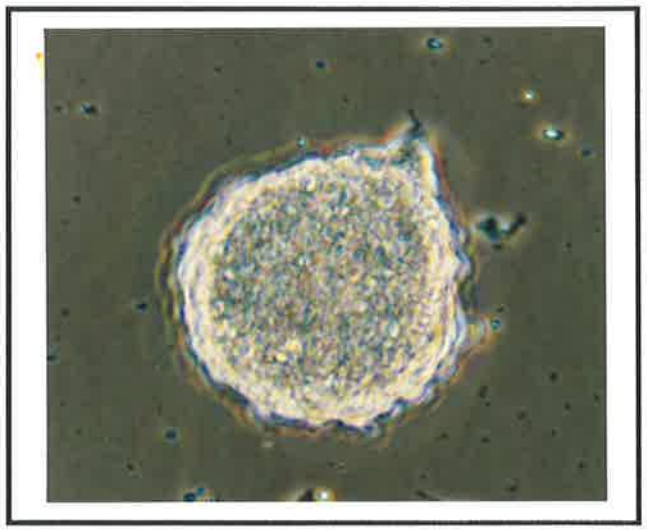
A



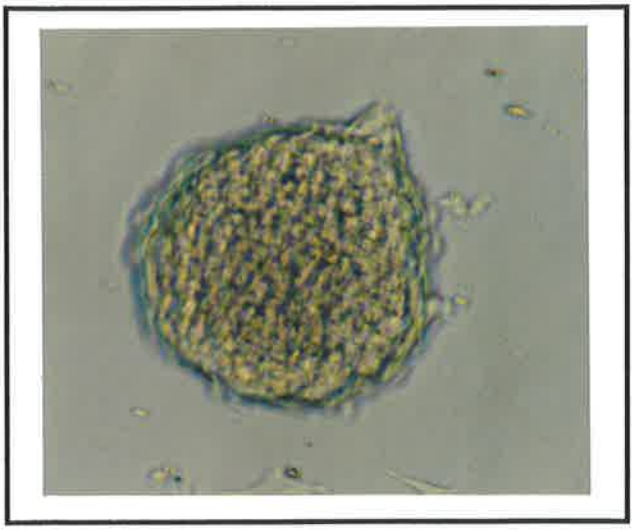
B



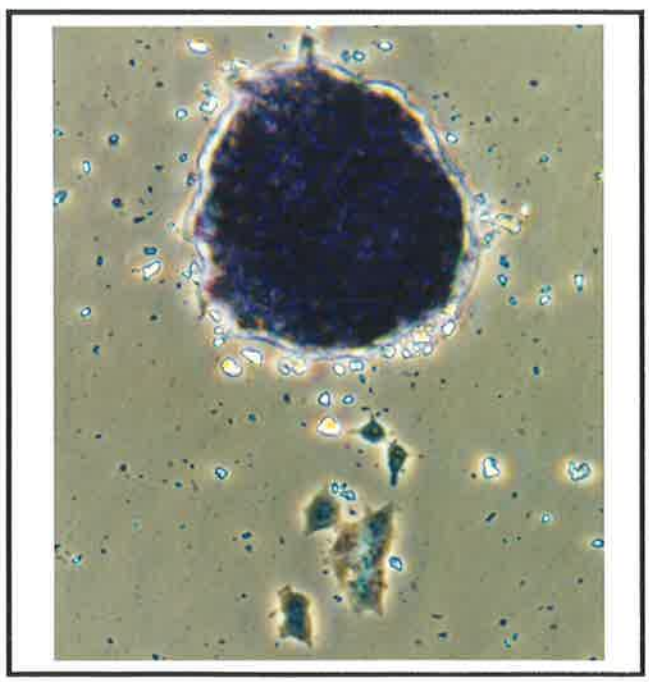
C



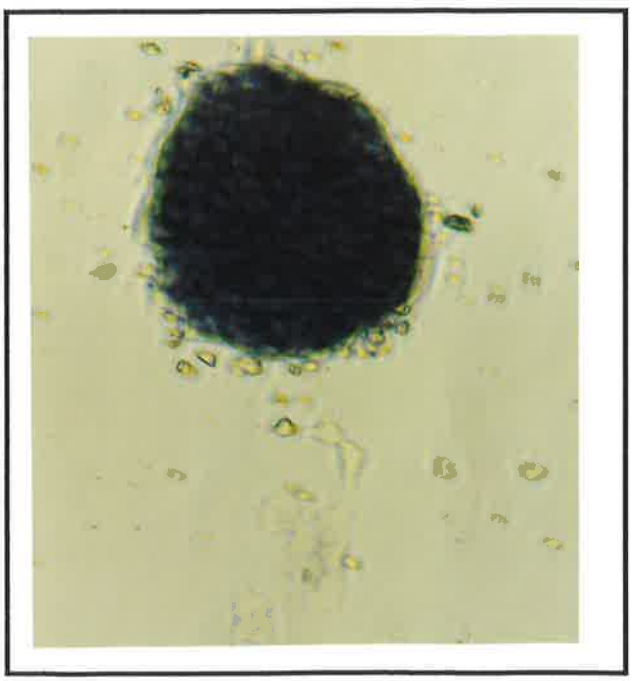
D



E



F



ES cell culture (Fig. 3.14E,F). This was consistent with the northern blot analysis of *MmoxA/B* expression (Fig. 3.11) which indicated that *MmoxA/B* was downregulated during ES cell differentiation.

Oct-4 expression in MBA induced ES cells was detected in nests of closely packed cells (Fig. 3.15A,B). These cells which expressed *Oct-4* represent the residual pluripotent cells which remain after ES cell induction. *Oct-4* was expressed by all of the cells within the nest, including the flatter cells at the periphery. As expected, a range of differentiated cells, as defined by their morphology, did not express *Oct-4* (Fig. 3.15A,B).

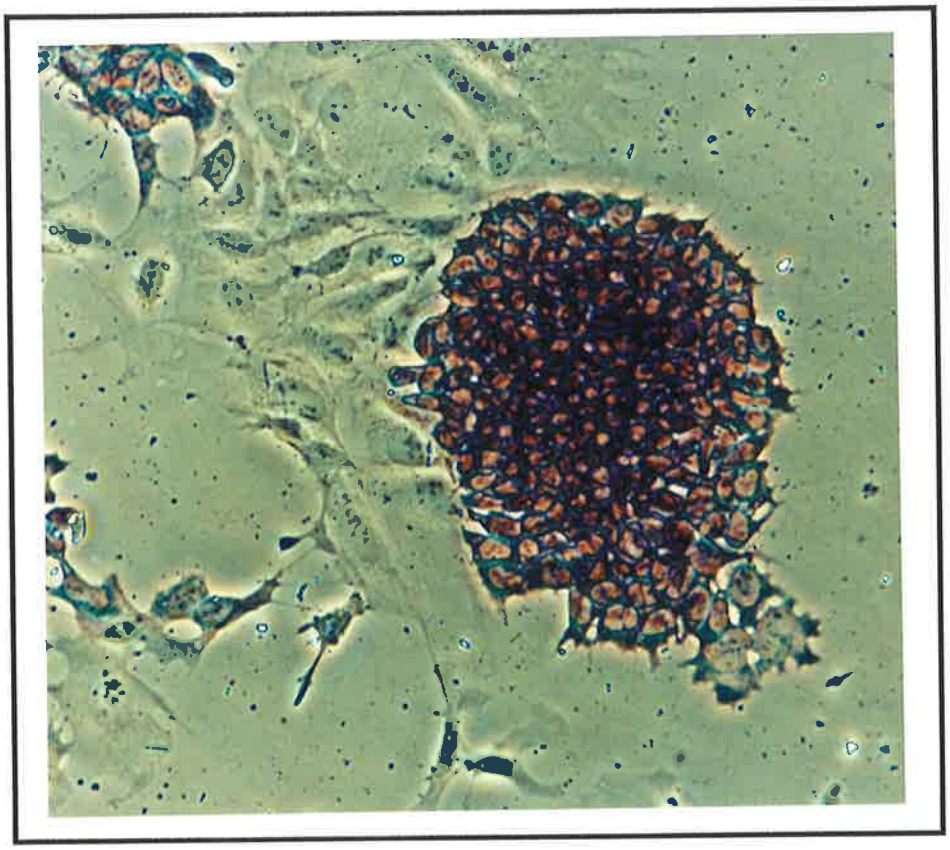
Like *Oct-4*, *MmoxA/B* was not expressed by the terminally differentiated ES cells induced by exposure to MBA (Fig 3.15C,D). This was consistent with the northern blot analysis of *MmoxA/B* expression which indicated that expression was downregulated during ES cell differentiation (Fig. 3.11).

MmoxA/B expression was detected in closely packed groups of cells, which presumably represented the residual pluripotent cells which remain after ES cell induction (3.15C,D). However, *MmoxA/B* differed from *Oct-4* in that *MmoxA/B* expression appeared to be downregulated in the cells at the periphery of the cell nest. Comparison between the peripheral nest cells in Figures 3.15A and 3.15C, indicated that these cells were of similar morphology and therefore probably represented common cell types. This *in situ* expression pattern suggested, therefore, that the nests of *Oct-4* positive cells harboured two distinct pluripotent cell populations: (central) *Oct-4*$^{+}$/*MmoxA/B*$^{+}$ cells and (peripheral) *Oct-4*$^{+}$/*MmoxA/B*$^{-}$ cells. This was consistent with the comparison of *MmoxA/B* and *Oct-4* expression by northern blot analysis which predicted the existence of distinct pluripotent cell populations which were discriminated by *MmoxA/B* expression. *In situ* and northern blot expression analysis therefore support a developmental role for *MMoxA* and/or *MMoxB* during early murine embryogenesis, perhaps by defining distinct differentiation intermediates which arise during pluripotent cell differentiation.

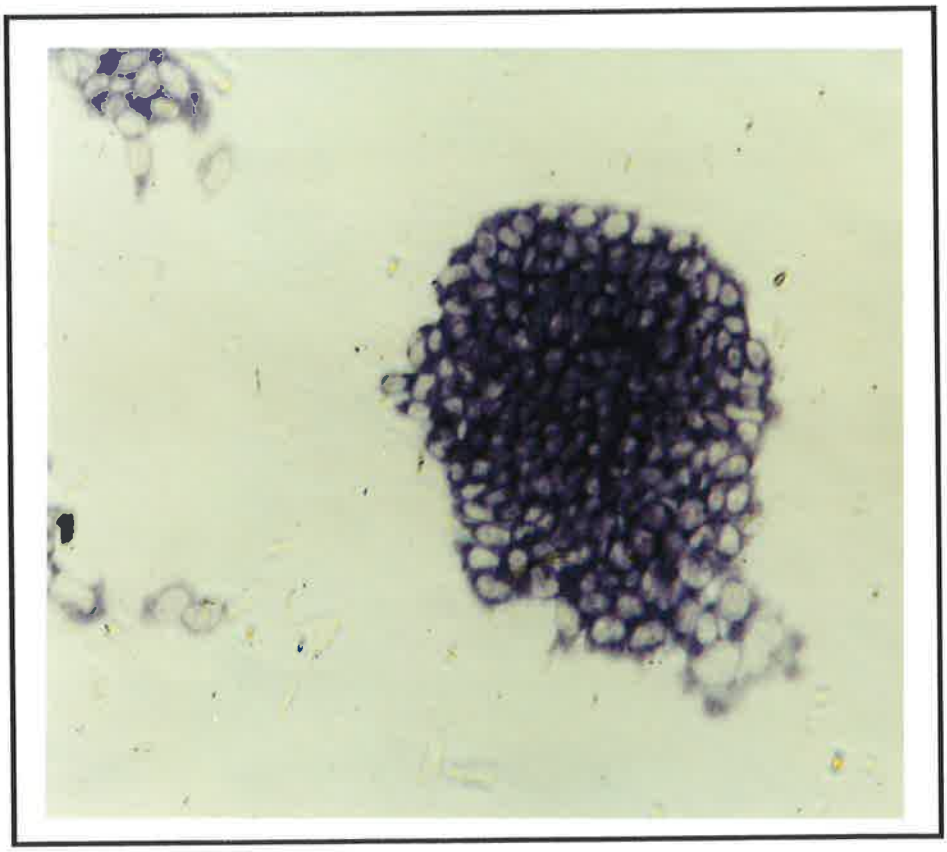
Figure 3.15 *In situ* analysis of *MmoxA/B* and *Oct-4* expression in MBA-induced ES cells.

Oct-4 expression (A, B) was detected using a 394 bp DIG labelled antisense riboprobe (2.3.25). *MmoxA/B* expression (C, D) was detected using a DIG labelled antisense riboprobe derived from the *MMoxB* PCR product (3.2.2). ES cells were induced with•MBA as described in section 2.4.2. The *Oct-4* and *MmoxA/B* plates were developed for 16 hours and 48 hours, respectively. Cells which express *Oct-4* or *MmoxA/B* are indicated by the purple staining. Photographs were taken under phase contrast (A, C) and brightfield illumination (B, D) at 50 X magnification.

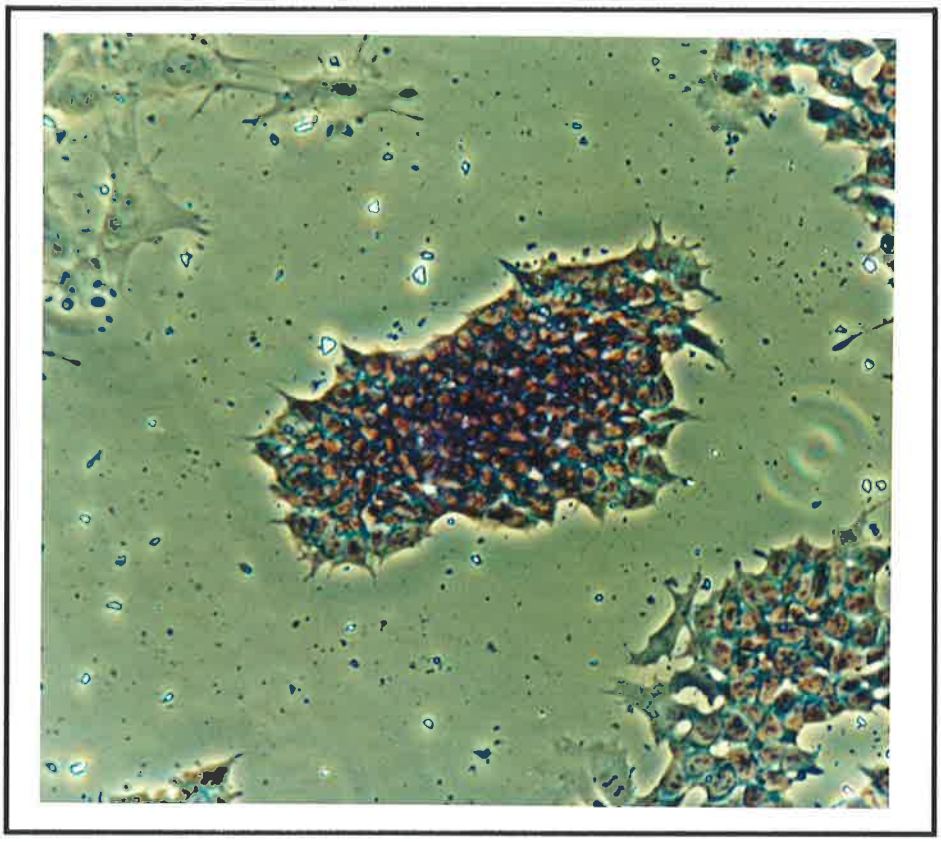
A



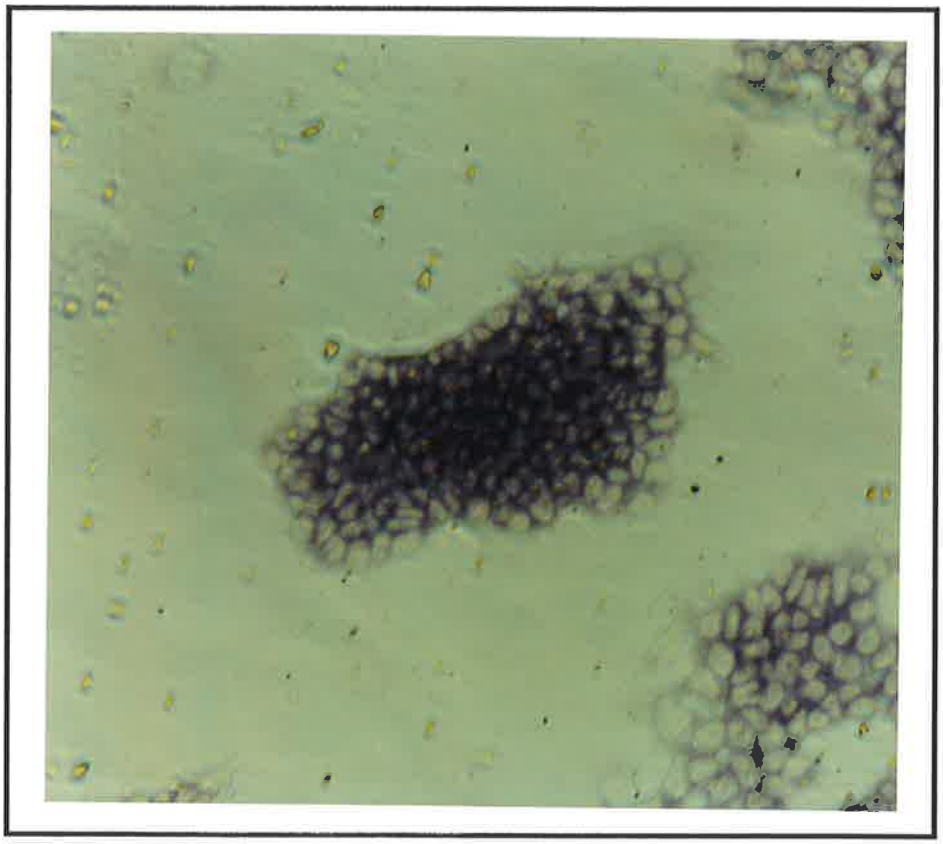
B



C



D



3.5 Discussion

A total of 13 homeobox gene sequences were isolated from ES cell cDNA using PCR primers based on conserved sequences in the *Antennapedia* and *Engrailed* homeodomain classes. Only five of the homeobox sequences which were isolated belong to the *Antennapedia* and *Engrailed* sequence classes i.e. *Hox-a7*, *Hox-a1*, *Hox-b1* and *Gsh-1* (*Antennapedia* class) and *En-2* (*Engrailed* class). The isolation of the remaining genes, which covered a diverse range of sequence classes, indicated that the PCR primers (especially 5'mANT' and 5'mEN') were capable of amplifying a significant proportion of all homeobox gene sequences. It is likely, therefore, that this PCR screen has identified many of the homeobox genes expressed by ES cells. However, the possibility that ES cells express additional homeobox genes which have not been isolated in this screen cannot be excluded.

Of the thirteen homeobox genes which were isolated from ES cell cDNA, *Hesx1*, *Hesx2*, *Hesx4*, *Hex*, *MMoxA* and *MMoxB* encoded relatively diverged partial homeodomain sequences. All of these genes, apart from *Hesx2*, were subsequently shown to be expressed by undifferentiated ES cells. Given the established role of homeobox genes in the regulation of cellular differentiation and the determination of cellular identity (1.7), the expression of homeobox genes with diverged homeodomain sequences in ES cells may reflect the unique cellular potential of pluripotent cells. In other words, pluripotent cells may express a distinct repertoire of developmental regulators, including homeobox genes, which reflects their ability to pursue a diverse range of cell lineages.

Of the genes expressed by ES cells, *Hesx1*, *Hesx4*, *Hex* and *MmoxA/B* represent the only known mammalian members of novel homeobox gene classes. A recurring theme of homeobox gene expression during embryogenesis, is the complementary and spatially restricted expression of homeobox genes encoding highly related homeodomain sequences, such as *Otx1* and *Otx2*, and *Emx1* and *Emx2* expression in the forebrain (Simeone *et al.*, 1992). This raises the possibility that additional members of these novel sequence classes may also be expressed by (sup-populations of) pluripotent cells. This possibility may be

investigated by PCR screening of ES cell cDNA with PCR primers specific for these novel sequence classes.

Previous analysis of homeobox genes expression during pluripotent cell differentiation has focused mainly on the *Hox* genes. Most *Hox* gene transcripts are not expressed at detectable levels in undifferentiated stem cells (Mavilio *et al* , 1988; Papalopulu *et al.*, 1990) and many *Hox* genes are induced within hours of RA treatment (Colberg-Poley *et al.*, 1985a, 1985b; Boncinelli *et al.*, 1991; Papalopulu *et al.*, 1990). Homeobox genes which are expressed at higher levels in undifferentiated ES cells than differentiated derivatives are rare, and are basically confined to the POU domain class genes *Oct-4* (Scholer *et al.*, 1990a; Rosner *et al.*, 1990; Yeom *et al.*, 1991; Okamoto *et al.*, 1990) and *Oct-6* (Suzuki *et al.*, 1990; Meijer *et al.*, 1990). Although, a developmental function for *Oct-6* has not been described, preliminary analysis of *Oct-4* null mutant embryos indicate that they die before implantation due to degeneration of the inner cell mass (Austin Smith, personal communication). The downregulation of *MMoxA* and/or *MMoxB*, *Hex*, *Hesx4* and *Hesx1* during ES cell differentiation, therefore suggests that these genes belong to a relatively small group of homeobox genes with this expression pattern, and are likely to have a developmental role in pluripotent cells during early murine embryogenesis. In addition, the differential regulation of the genes in response to different ES cell inducers, suggests that *MMoxA* and/or *MMoxB*, *Hex*, *Hesx4* and *Hesx1* may have distinct developmental roles. The expression of *Hex* and *Hesx4* in the terminally differentiated RA-induced cells, indicates that these genes may also be expressed (and have a developmental role) during latter stages of ontogeny.

Comparison of *MmoxA/B* and *Oct-4* expression in undifferentiated ES cells and in X-1 and X+3 cells, suggested the existence of two distinct populations of pluripotent cells i.e. *MmoxA/B*$^{+}$/*Oct-4*$^{+}$ (undifferentiated ES cells) and *MmoxA/B*$^{-}$/*Oct-4*$^{+}$ (X-1 and X+3 cells). The level of *MmoxA/B* expression in the X+1 cells was approximately 50% of the level in undifferentiated ES cells. Despite the identical morphology of the X+1, X-1 and X+3 cells, the reduced *MmoxA/B* expression in the X+1 cells suggested that these cells are an "intermediate" in the transformation of ES cells to X cells. This was consistent with

previous comparisons of marker gene expression in ES cells, X+1 and X+3 cells (Bettess, 1993). The reduced levels of *MmoxA/B* in the X+1 cells may have been due to the presence of DIA/LIF in culture medium, which is able to inhibit ES cell differentiation and may therefore "slow" the ES to X cell transition.

Tissue culture and biochemical analysis of X cells has indicated that these cells are pluripotent, and may represent an *in vitro* equivalent of the primitive ectoderm (see 3.4.1). Although the pluripotent inner cell mass and primitive ectoderm cells are morphologically distinct, genes which expressed differentially between these tissues have not been described. Therefore, the greatly reduced expression of *MmoxA/B* levels in the X-1 and the X+3 cells compared with undifferentiated ES cells, suggests that *MMoxA* and/or *MMoxB* may be downregulated during formation of the primitive ectoderm from the inner cell mass. Given that *MMoxA* and *MMoxB* are homeobox genes, it is therefore possible that the differential expression of *MMoxA* and/or *MMoxB* may be one of the key differences in the molecular components which establish cellular identity in the inner cell mass versus the primitive ectoderm during the early stages of embryogenesis. *In situ* analysis of *MMoxA* and *MMoxB* expression in the early embryo is required to determine whether these genes are differentially expressed in the inner cell mass and primitive ectoderm.

Comparison of *MmoxA/B* and *Oct-4* expression in uninduced ES cells and spontaneously and chemically induced ES cells also suggested the existence of two distinct populations of pluripotent cells (as defined by *Oct-4* expression), which were discriminated by their expression of *MmoxA/B*. This suggests that during ES cell differentiation, *MmoxA/B* expression is downregulated before *Oct-4*, to generate a "differentiation intermediate" with the phenotype *MmoxA/B*$^-$/*Oct-4*$^+$. As X-1 and X+3 cells also shared this phenotype, it is possible that the proposed "differentiation intermediate" and the X-1 and X+3 cells may represent equivalent cell populations. Furthermore, given that X cells are believed to be an *in vitro* equivalent of the primitive ectoderm, the possibility that ES cell differentiation may progress via a primitive ectoderm-like intermediate *in vitro*, is consistent with transformation of the inner cell mass into the primitive ectoderm as an intermediate of pluripotent cell differentiation *in vivo*.

In situ analysis of ES cell monolayers confirmed that *MmoxA/B* was expressed by undifferentiated ES cells, and identified residual stem cells which remained after MBA induction which express *MmoxA/B*. In comparison to *Oct-4*, *MmoxA/B* expression appeared to be restricted to a subset of the residual stem cells, which supported the existence of the proposed *MmoxA/B*$^{+}$/*Oct-4*$^{+}$ and *MMoxA/B*$^{-}$/*Oct-4*$^{+}$ cell types. Further comparison of *MmoxA/B* and *Oct-4* expression at the cellular level would be facilitated by the development of a double staining technique which would allow visualization of *MmoxA/B* and *Oct-4* expression simultaneously.

Relative levels of *Hex*, *Hesx4* and *Oct-4* expression in MBA- and DMSO-induced ES cells indicated that *Hex* and *Hesx4,* like *MmoxA/B*, were downregulated during pluripotent cell differentiation using these regimes. However, *Hex* and *Hesx4* expression differed from *MmoxA/B* in two respects. Firstly, *Hex* and *Hesx4* expression was not downregulated in X cells, suggesting that these genes are likely to be expressed by the primitive ectoderm during embryogenesis. Secondly, *Hex* and *Hesx4* were expressed by the spontaneously and RA-induced ES cells. This may indicate expression of these genes during latter stages of embryogenesis, and is consistent with the expression of *Hex* in differentiated cell types (Bedford *et al.*, 1993).

The greater relative downregulation of *Hex* and *Hesx4* expression compared with *Oct-4* in the MBA- and DMSO-induced ES cells suggested that these genes may define a pluripotent cell intermediate of the phenotype *Hex*$^{-}$/*Hesx4*$^{-}$/*Oct-4*$^{+}$. Alternatively, *Hex* and *Hesx4* may be down-regulated independently during ES cell differentiation, and define distinct pluripotent differentiation intermediates. Furthermore, as *Hex* and *Hesx4* were not downregulated in X cells (unlike *MmoxA/B*), it is likely that *Hex* and *Hesx4* would define a pluripotent cell intermediate which is different from the *MmoxA/B*$^{-}$/*Oct-4*$^{+}$ cells. This would support distinct developmental roles for *Hex* and *Hesx4* versus *MMoxA* and/or *MMoxB* during pluripotent cell differentiation *in vivo*.

Finally, this investigation has established for the first time that ES cells express a range of homeobox genes. This indicates that this class of transcription factors, which control developmental decisions at latter stages of embryogenesis, are likely to also play a

role during early development. Furthermore, homeobox genes expressed by ES cells were regulated differently in response to a range of differentiation stimuli suggesting that these genes have distinct biological functions during pluripotent cell differentiation.

CHAPTER 4: ISOLATION OF *Hesx1* cDNA CLONES AND ANALYSIS OF *Hesx1* EXPRESSION

CHAPTER 4: ISOLATION OF *Hesx1* cDNA CLONES AND ANALYSIS OF *Hesx1* EXPRESSION

4.1 Introduction

Hesx1 was the first of three novel homeobox genes isolated from ES cells using PCR, and was selected for further analysis on the basis of two interesting features. Firstly, comparison between *Hesx1* and *Oct-4* expression in undifferentiated and induced ES cells indicated that *Hesx1* expression was downregulated during ES cell differentiation, particularly in response to retinoic acid induction. This expression pattern *in vitro* suggested that *Hesx1* is expressed in the pluripotent cell lineages during early embryogenesis, and may play a role in the regulation of stem cell renewal and differentiation in the early murine embryo. Secondly, analysis of the partial *Hesx1* homeobox sequence indicated that *Hesx1*, along with the *Xenopus* gene XANF-1, defines a novel class of homeobox gene sequences. Further *Hesx1* sequence was required to compare the complete *Hesx1* homeobox sequence with XANF-1, and to determine whether these related homeobox genes shared more extensive homology.

Northern blot analysis indicated that *Hesx1* expression levels were low (even in ES cells) and the relative expression of the alternative transcripts was difficult to resolve. More extensive analysis of *Hesx1* expression *in vitro* would require an assay of greater sensitivity which could easily discriminate between the *Hesx1* transcripts. Isolation of *Hesx1* cDNA clones was essential for the development of this assay.

This chapter describes the isolation and characterization of *Hesx1* cDNA clones corresponding to the 1.0 kb and 1.2 kb *Hesx1* transcripts. RNAase protection analysis using antisense riboprobes generated from one of these clones allowed subsequent analysis of *Hesx1* expression in undifferentiated and differentiated ES cells, and in tissue specific RNA samples derived from the murine embryo. The work described in this chapter has been published (Thomas and Rathjen, 1992; Thomas *et al.*, 1994).

4.2 Isolation, Sequencing and Analysis of *Hesx1* cDNA Clones

4.2.1 Isolation of *Hesx1* cDNA Clones

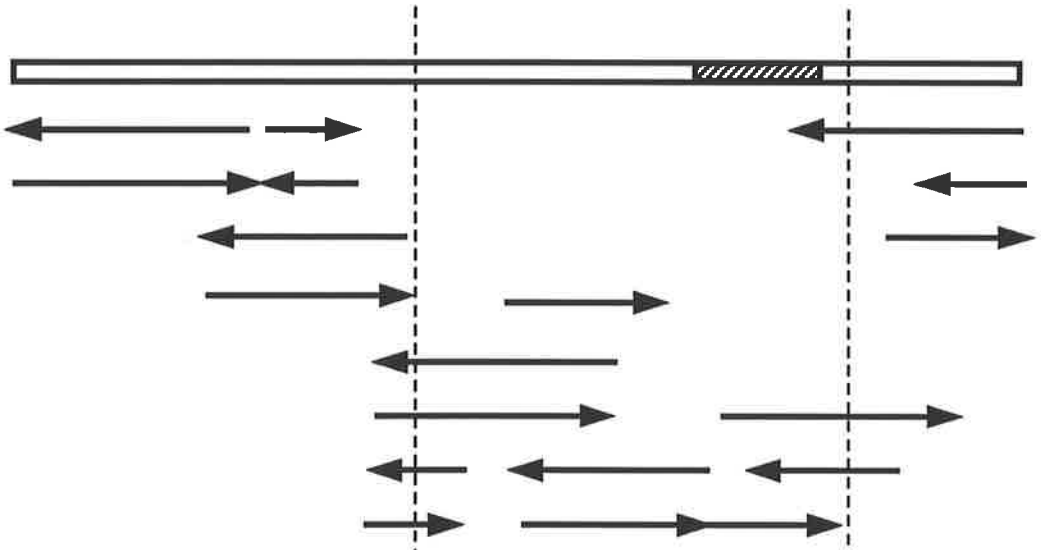
A Lambda Zap II cDNA library derived from undifferentiated D3 ES cells (Doetschmann *et al.*, 1985) maintained in the presence of DIA/LIF was screened using the *Hesx1* partial homeobox sequence which included the flanking primers En5 and En3 (3.2.1). A total of 1.25 X 10^6 plaques were screened and five third round duplicate positive plaques were isolated and Zapped into pBluescript. The cDNA inserts were released by digestion with *EcoRI* and sized by agarose gel electrophoresis. Four of the *Hesx1* cDNA clones contained inserts of 1427 bp and the fifth clone contained a 1509 bp insert. Given that the two *Hesx1* transcripts were 1.0 kb and 1.2 kb, it seemed likely that these cDNA clones contained additional sequence not derived from the *Hesx1* transcripts.

Approximately 200 bp of 5' and 3' terminal sequence from each of the five clones was determined which indicated that the four clones containing 1427 bp inserts were identical (data not shown). The 1427 bp and 1509 bp inserts were digested with *Alu*I, *Hae*III or *Rsa*I and sub-cloned into pT7T319U. This generated a series of overlapping clones which were sequenced from both strands (Fig. 4.1). Sequence analysis revealed that the cDNA inserts contained 625 bp of common sequence in the central portion of each clone, and did not share any sequence homology 3' or 5' of this common region (Fig. 4.1). Comparison of the 1427 bp cDNA sequence with the GenBank database indicated that the 563 bp 5' unique region shared greater than 90% homology with the rabbit eukaryotic release factor (eRF) cDNA sequence. This suggested that cDNA derived from this gene had ligated to the 5' end of the *Hesx1* cDNA during construction of the library. The artifactual nature of this 5' sequence was confirmed by probing an ES cell northern blot with a probe derived from the 5' unique sequence. This analysis detected a single 3 kb band which could not be attributed to *Hesx1* due to its size (data not shown). The 5' unique end of the 1509 cDNA insert also appeared to be artifactual as a a 280 bp fragment derived from this region failed to hybridize to the *Hesx1* genomic clone (Hesx1g, Chapter 5). This was confirmed

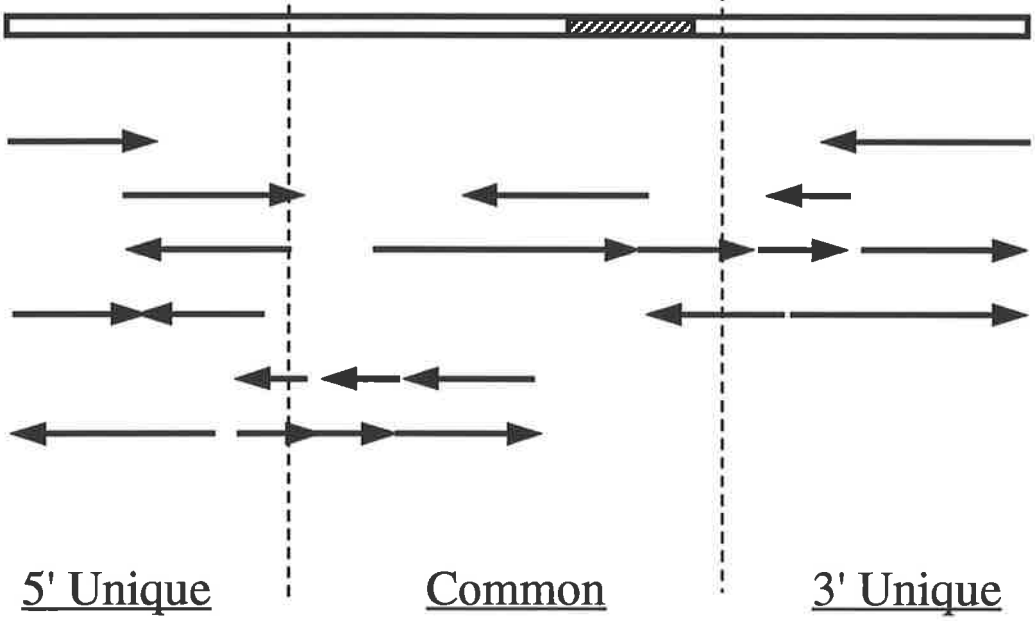
Figure 4.1 Sequencing strategy for the Hesx1a and Hesx1b cDNA clones.

The Hesx1a and Hesx1b cDNA clones are indicated by the open bars, and sequenced regions are shown by the arrows. The dotted lines separate the 5' Unique, Common and 3' Unique sequences. The striped box within the common region shows the location of the homeobox.

Hesx1a



Hesx1b



5' Unique Common 3' Unique

200 bp

by direct sequence comparison of the cDNA and genomic clones which failed to detect any sequence identity across the 5' 392 bp of the cDNA clone.

The *bona fide Hesx1* sequence derived from the cDNA clones (as determined by comparison with the *Hesx1* genomic clone (5.2.2)) is shown in Figure 4.2. The Hesx1a clone encoded 847 bp of *Hesx1* cDNA, which included a poly A tail, and the Hesx1b cDNA clone contained 1117 bp of *Hesx1* sequence. The isolation of two distinct *Hesx1* cDNA clones was consistent with the previous detection by northern blot analysis of two *Hesx1* transcripts of 1.2 kb and 1.0 kb (Chapter 3). Hesx1a and Hesx1b were identical in sequence from position -65 (the 5' end of Hesx1a) to position 560 bp, but differed in sequence downstream of this point. This indicated that the 1.2 kb and 1.0 kb *Hesx1* transcripts may be 5' coterminal. An additional 13 bases of sequence at the 5' end of the Hesx1b clone (-78 to -66) were identified by sequence comparison with the *Hesx1* genomic clone (5.2.2).

4.2.2 Sequence Analysis of the Hesx1a and Hesx1b cDNA Clones

The 180 bp *Hesx1* homeobox spanning positions 322 to 501 was identified within the Hesx1a/Hesx1b common sequence (Fig. 4.2A). The homeobox sequence between positions 345 and 453 was identical to the partial *Hesx1* homeobox sequence identified by PCR screening of ES cell cDNA (3.2.1).

A dinucleotide CA repeat sequence containing 12 direct dinucleotide repeats was located within the unique 3' region of the Hesx1b sequence (Fig. 4.2B). $(CA)_n$ repeats are highly polymorphic with respect to the number of dinucleotide repeats, particularly if n is greater than ten (Weber, 1990), indicating that the *Hesx1* (CA) repeat is likely to be polymorphic. The biological function of $(CA)_n$ repeats is not known (Stallings *et al.*, 1991).

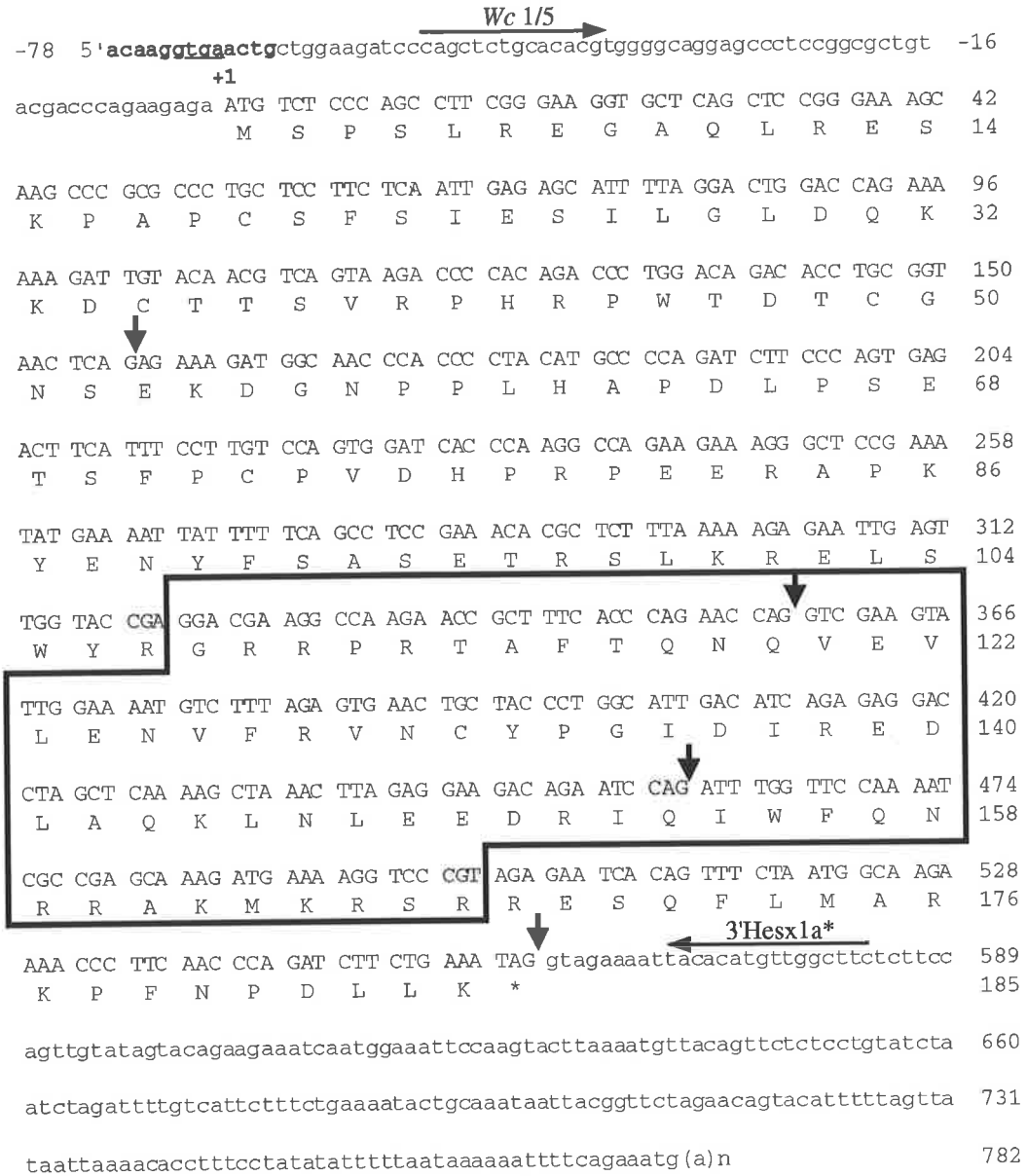
4.2.3 Identification and Analysis of the *Hesx1* Open Reading Frame

An identical 185 amino acid open reading frame, containing the *Hesx1* homeodomain (amino acids 108-167) was identified in the Hesx1a and Hesx1b cDNA clone sequences (Fig. 4.2A). The identity of the initiation codon for the Hesx1b open reading frame was

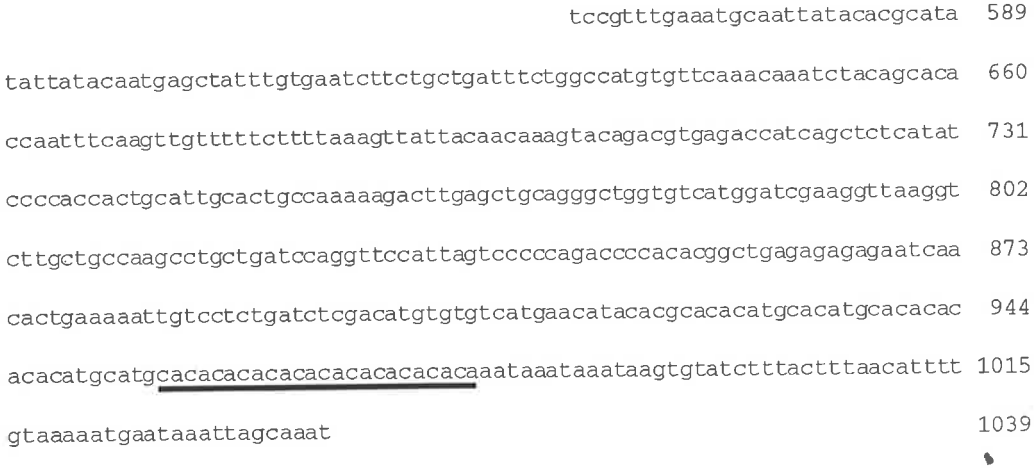
Figure 4.2 Nucleotide sequence of the Hesx1a (A) and Hesx1b (B) cDNA clones.

Nucleotide sequence common to the Hesx1a and Hesx1b clones spans nucleotide positions -65 to 560 and encodes a 185 amino acid open reading frame from position 1 to 558. The additional 13 nucleotides at the 5' end of the Hesx1b clone which contain an in frame stop codon (underlined) are shown in bold. Nucleotide sequence within the boxed region corresponds to the homeobox. Positions of introns within the genomic clone (5.2.2) are indicated by arrows. The 3' unique sequences of the Hesx1b clone, which correspond to nucleotides downstream of position 560 in Hesx1a, are shown in B. The positions of the The 12 consecutive CA repeats in the Hesx1b transcript are indicated by a bold underline. The positions of the 3Hesx1a and Wc 1/5 primers are also shown.

A


Wc 1/5
-78 5'acaaggtgaactgctggaagatcccagctctgcacacgtggggcaggagccctccggcgctgt -16
+1
acgacccagaagaga ATG TCT CCC AGC CTT CGG GAA GGT GCT CAG CTC CGG GAA AGC 42
M S P S L R E G A Q L R E S 14
AAG CCC GCG CCC TGC TCC TTC TCA ATT GAG AGC ATT TTA GGA CTG GAC CAG AAA 96
K P A P C S F S I E S I L G L D Q K 32
AAA GAT TGT ACA ACG TCA GTA AGA CCC CAC AGA CCC TGG ACA GAC ACC TGC GGT 150
K D C T T S V R P H R P W T D T C G 50
AAC TCA GAG AAA GAT GGC AAC CCA CCC CTA CAT GCC CCA GAT CTT CCC AGT GAG 204
N S E K D G N P P L H A P D L P S E 68
ACT TCA TTT CCT TGT CCA GTG GAT CAC CCA AGG CCA GAA GAA AGG GCT CCG AAA 258
T S F P C P V D H P R P E E R A P K 86
TAT GAA AAT TAT TTT TCA GCC TCC GAA ACA CGC TCT TTA AAA AGA GAA TTG AGT 312
Y E N Y F S A S E T R S L K R E L S 104
TGG TAC CGA GGA CGA AGG CCA AGA ACC GCT TTC ACC CAG AAC CAG GTC GAA GTA 366
W Y R G R R P R T A F T Q N Q V E V 122
TTG GAA AAT GTC TTT AGA GTG AAC TGC TAC CCT GGC ATT GAC ATC AGA GAG GAC 420
L E N V F R V N C Y P G I D I R E D 140
CTA GCT CAA AAG CTA AAC TTA GAG GAA GAC AGA ATC CAG ATT TGG TTC CAA AAT 474
L A Q K L N L E E D R I Q I W F Q N 158
CGC CGA GCA AAG ATG AAA AGG TCC CGT AGA GAA TCA CAG TTT CTA ATG GCA AGA 528
R R A K M K R S R R E S Q F L M A R 176
3'Hesx1a*
AAA CCC TTC AAC CCA GAT CTT CTG AAA TAG gtagaaaattacacatgttggcttctcttcc 589
K P F N P D L L K * 185
agttgtatagtacagaagaaatcaatggaaattccaagtacttaaaatgttacagttctctcctgtatcta 660
atctagattttgtcattctttctgaaaatactgcaaataattacggttctagaacagtacatttttagtta 731
taattaaaacacctttcctatatatttttaataaaaaattttcagaaatg(a)n 782

B


tccgtttgaaatgcaattatacacgcata 589
tattatacaatgagctatttgtgaatcttctgctgatttctggccatgtgttcaaacaaatctacagcaca 660
ccaatttcaagttgtttttctttaaagttattacaacaaagtacagacgtgagaccatcagctctcatat 731
ccccaccactgcattgcactgccaaaaagacttgagctgcagggctggtgtcatggatcgaaggttaaggt 802
cttgctgccaagcctgctgatccaggttccattagtcccccagaccccacacggctgagagagagaatcaa 873
cactgaaaaattgtcctctgatctcgacatgtgtgtcatgaacatacacgcacacatgcacatgcacacac 944
acacatgcatgcacacacacacacacacacacacaaataaataaataagtgtatctttactttaacatttt 1015
gtaaaaatgaataaattagcaaat 1039

confirmed by the presence of an upstream in-frame stop codon spanning positions -72 to -70 in the Hesx1b cDNA (underlined in Fig. 4.2A). Insufficient 5' sequence was available to determine whether this in frame upstream stop codon was also present in the Hesx1a cDNA. The termination codon for the open reading frame was located immediately upstream of the sequence divergence between the *Hesx1* cDNAs, indicating that the alternative transcripts encoded identical proteins and differed only in the 3' untranslated region.

The *Hesx1* homeodomain was identified in the carboxy-terminal half of the putative *Hesx1* protein, spanning positions 108-167 of the open reading frame. No additional conserved protein motifs were detected within the *Hesx1* open reading frame using the DNasis software package.

4.2.4 Sequence Comparison of the *Hesx1* Homeodomain

Sequence comparison showed that the entire *Hesx1* homeodomain shared 80% identity with the *Xenopus* homeobox gene XANF-1 (Zariasky *et al.*, 1992) but was not >50% related to other homeodomain sequences (Fig. 4.3). This identifies *Hesx1* and XANF-1 as founding members of a novel homeodomain class as had previously been suggested from analysis of partial homeodomain sequences (3.2.3).

Comparison between the entire *Hesx1* and partial XANF-1 open reading frames revealed additional structural similarities (Fig. 4.4). *Hesx1* and XANF-1 were identical at the 8 upstream and 6 downstream amino acids which flank the homeodomain, and terminated 15 amino acids downstream of the homeodomain. Apart from the sequence identity within and flanking the homeodomain, a stretch of 34 amino acid residues at the amino terminus of the *Hesx1* open reading frame shared 56% homology with XANF-1. Little homology was detected in the remaining regions of the *Hesx1* and XANF-1 primary sequences. The similarities in primary structure suggest that the *Hesx1* and XANF-1 genes may have evolved from a common ancestral gene, but are probably not sufficient to consider *Hesx1* and XANF-1 as sequence homologs.

Figure 4.3 Sequence comparison of the *Hesx1* homeodomain.

The positions of the three alpha helices and homeodomain consensus sequence (Scott *et al.*, 1989) are shown at the top of the figure. Consensus *Hesx1* residues are shown in bold. Amino acid residues which are conserved in the *Hesx1*, XANF-1 (Zariasky *et al.*, 1992) and *Antennapedia* (Scott and Weiner, 1984a; McGinnis *et al.*, 1984) homeodomains are indicated by dashes. The percentage homology with *Hesx1* is shown on the right.

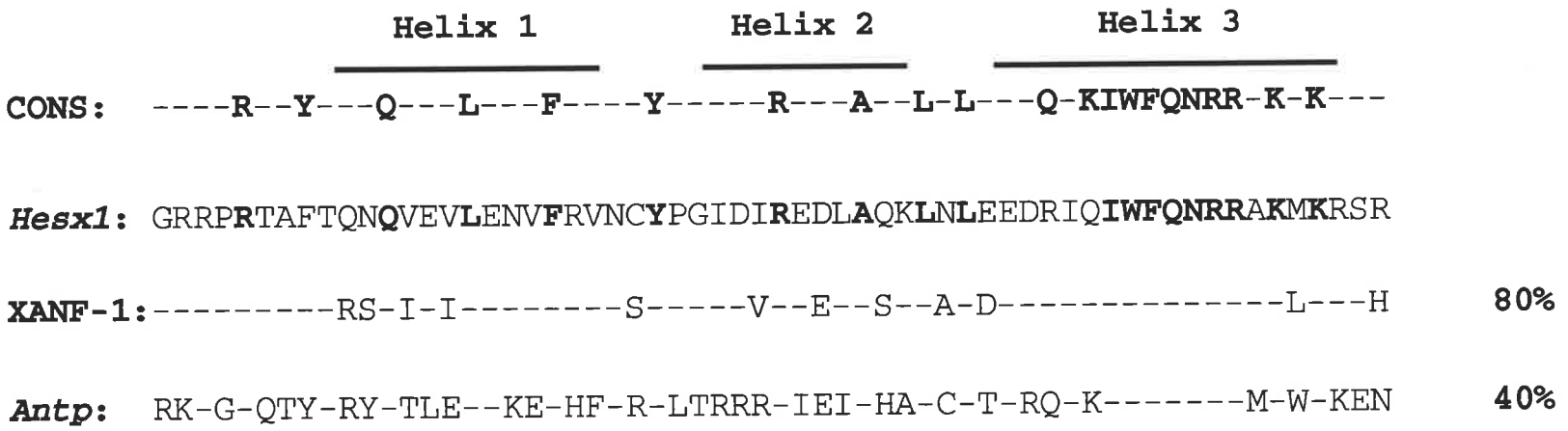

Helix 1
Helix 2
Helix 3
CONS: ----R--Y---Q---L---F----Y-----R---A--L-L---Q-KIWFQNRR-K-K---
Hesx1: GRRPRTAFTQNQVEVLENVFRVNCYPGIDIREDLAQKLNLEEDRIQIWFQNRRAKMKRSR
XANF-1:---------RS-I-I--------S-----V--E--S--A-D--------------L---H
80%
Antp: RK-G-QTY-RY-TLE--KE-HF-R-LTRRR-IEI-HA-C-T-RQ-K-------M-W-KEN
40%

Figure 4.4 Comparison of the *Hesx1* and XANF-1 open reading frames.

Positions of identity are indicated by vertical lines and homeodomain sequences are shown in boldface. Numbers on the right refer to amino acid positions in the *Hesx1* (Fig. 4.2) and XANF-1 (Zariasky *et al.*, 1992) open reading frames.

```
XANF-1                                    KSQLSLTSLT  10

Hesx1  MSPSLREGAQLRESKPAPCSFSIESILGLDQKKDCTTS  38
       ||| |  |  | |    | | ||| ||||| | |
XANF-1 MSPALQKGSSLMENRSPPSSSSIEHILGLDKKTDWLHH  48

Hesx1  VRPHRPWTDTCGNSEKDGNPPLHAPDLPSETSFPC-PV  75
                |              |             |
XANF-1 PLLSTTAVIECSSKGVVNGTCWQIPVIACDLPIQVHAV  86

Hesx1  DHPRPEERAPKYENYFSASETRSLKRELSWYRGRRPRT  113
            ||     |  |        ||||||||||||||
XANF-1 HRSEEEETKIRLEKCFGDEDRLTYKRELSWYRGRRPRT  124

Hesx1  AFTQNQVEVLENVFRVNCYPGIDIREDLAQKLNLEEDR  152
       |||  | | |||||||| ||||| || || || | |||
XANF-1 AFTRSQIEILENVFRVNSYPGIDVREELASKLALDEDR  162

Hesx1  IQIWFQNRRAKNKRSRRESQFLMARKPFNPDLLK*     185
       |||||||||||  |||  ||||||
XANF-1 IQIWFQNRRAKLKRSHRESQFLIVKDSLSSKIQE*     196
```

4.2.5 Derivation of the Hesx1a and Hesx1b Transcripts

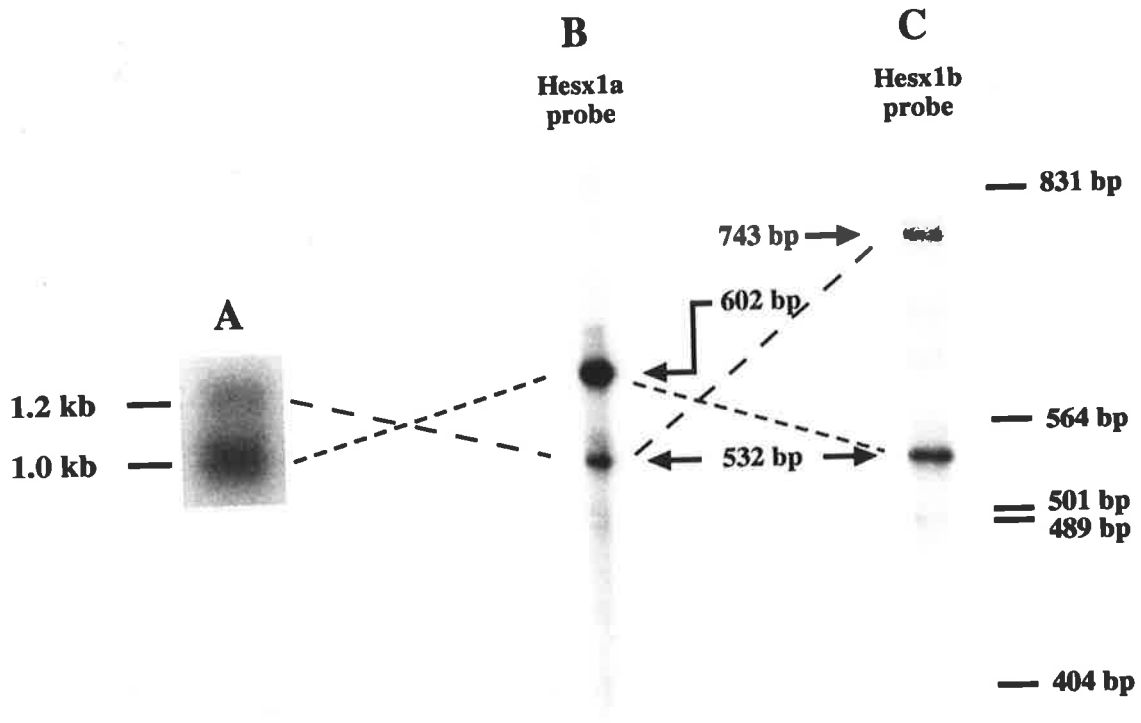
The relative sizes of the Hesx1a and Hesx1b clones suggested that they were derived from the 1.0 kb and 1.2 kb transcripts respectively. This was confirmed by comparison of RNAase protection products generated by antisense riboprobes derived from the 3' ends of the two cDNAs (Fig. 4.5). The most abundant *Hesx1* transcript expressed by undifferentiated ES cells was the 1.0 kb transcript (Fig. 4.5A). RNAase protection using an antisense riboprobe spanning positions 26 to 628 of the Hesx1a clone (Fig. 4.5B,D) generated a more intense 602 bp band (resulting from protection across the entire probe), and a less intense 532 bp band (resulting from protection across the common cDNA sequence). The greater relative intensity of the 602 bp band indicated that the Hesx1a clone was derived from the more abundant 1.0 kb transcript. The reciprocal experiment, using a Hesx1b antisense probe spanning positions 26 to 770 (Fig. 4.5D), was also carried out (Fig. 4.5C). The intensity of the 532 bp band (corresponding to the cDNA common region) was less than the 743 bp band (resulting from protection across the entire probe) indicating that the Hesx1b cDNA clone was derived from the less abundant 1.2 kb *Hesx1* transcript. The Hesx1a and Hesx1b cDNAs therefore correspond to the 1.0 kb and 1.2 kb *Hesx1* transcripts, respectively.

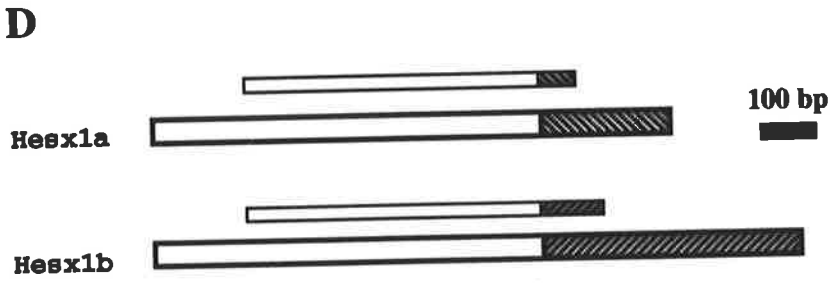
4.3 RNAase Protection Analysis of *Hesx1* Expression

Northern blot analysis of *Hesx1* expression, indicated that the 1.0 kb and 1.2 kb *Hesx1* transcripts were expressed at extremely low levels. In addition, the relative expression of the alternative *Hesx1* transcripts and was difficult to resolve using this technique. Therefore, further analysis of *Hesx1* expression in was carried out using the more sensitive RNAase protection assay (Krieg and Melton, 1987) with the Hesx1a derived riboprobe (4.2.5). The RNAase protection products of 532 bp and 602 bp corresponded to the 1.2 kb and 1.0 kb *Hesx1* transcripts respectively.

Figure 4.5 The Hesx1a and Hesx1b cDNAs are derived from the 1.0 kb and 1.2 kb *Hesx1* transcripts respectively.

A. Northern blot analysis of approximately 10 μg $(A)_n^+$ MBL-5 ES cell RNA probed with an antisense 402 bp *Hesx1* riboprobe derived from the common Hesx1a and Hesx1b sequences. The *Hesx1* antisense riboprobe spanned positions 31 to 432 of the Hesx1a cDNA. **B., C.** RNAase protection analysis of 20 μg undifferentiated MBL-5 ES cell RNA was carried out using antisense riboprobes spanning positions 26 to 628 and 26 to 770 of the Hesx1a (B) and Hesx1b (C) cDNAs respectively (see 2.3.21). *Hesx1* bands derived from a common transcript are connected by dashed lines. **D.** Molecular origin of the Hesx1a and Hesx1b riboprobes. Wide boxes represent *Hesx1* transcripts and corresponding riboprobes are shown as narrow boxes above. 3' unique sequences are indicated by diagonal shading.


B
Hesx1a
probe
C
Hesx1b
probe
831 bp
743 bp
602 bp
A
1.2 kb
1.0 kb
564 bp
532 bp
501 bp
489 bp
404 bp


D
100 bp
Hesx1a
Hesx1b

4.3.1 *Hesx1* Expression in ES cells and Differentiated Derivatives

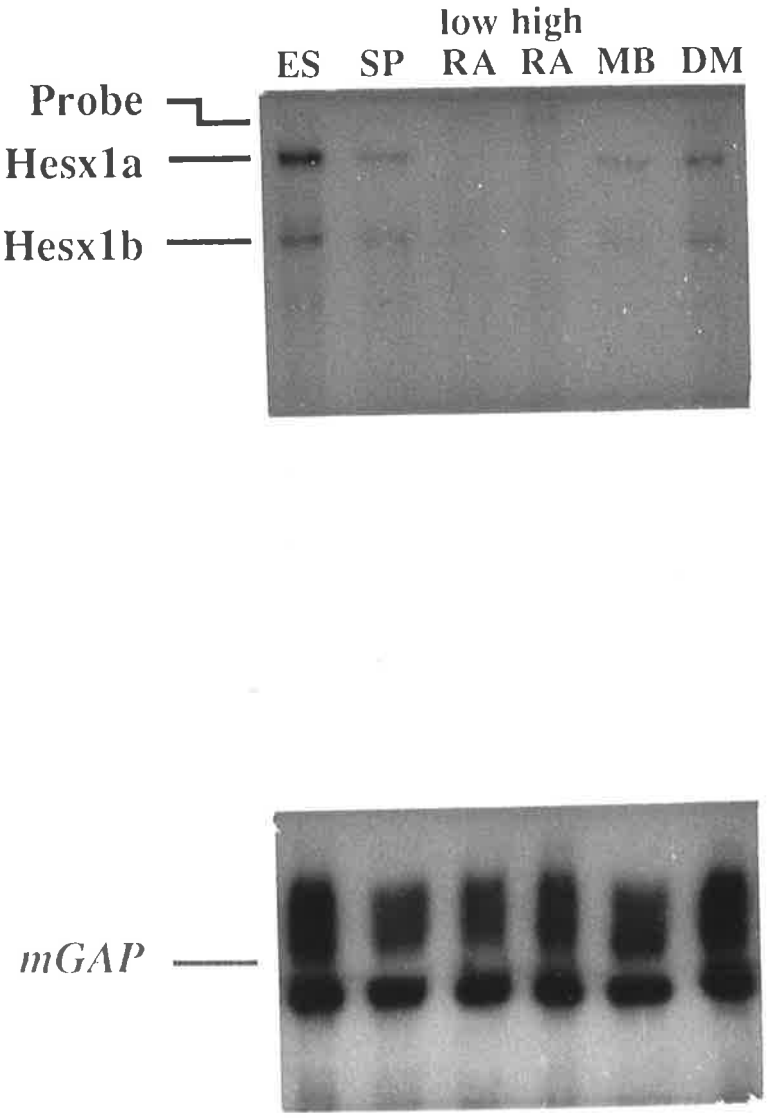
Comparison of *Hesx1* and *Oct-4* expression by northern blot analysis indicated that *Hesx1* was expressed by undifferentiated ES cells, and was downregulated during spontaneously- and chemically-induced ES cell differentiation (Chapter 3). Analysis of *Hesx1* expression in uninduced and induced ES cells was repeated using RNAase protection. The ES cells and induced ES cells analysed here and in Chapter 3 were generated independently.

20 μg of ES cell, X cell and chemically- and spontaneously-induced ES cell RNA was assayed for *Hesx1* expression using RNAase protection (Fig. 4.6). *Hesx1* expression was detected at highest levels in undifferentiated ES cells. Terminal differentiation of ES cells with 10 μM (high) or 2 μM (low) RA resulted in substantial downregulation of *Hesx1* expression. Reduced levels of *Hesx1* expression were detected in the spontaneous-, MBA- and DMSO-induced ES cells. These lower levels of *Hesx1* expression (compared with undifferentiated ES cells) may have been due to the residual undifferentiated ES cells which persisted within these cultures. This analysis therefore suggested that *Hesx1* was downregulated during ES cell differentiation by spontaneous and chemical induction, and was consistent with the northern blot analysis described in Chapter 3. Some variation in the relative *Hesx1* expression levels, as determined by northern blot and RNAase protection analysis, was detected in the spontaneously and chemically induced ES cells. This probably reflected variation in the level of residual pluripotent cells in these (independently generated) cultures.

In all samples in which *Hesx1* expression was detected, the 602 bp band was of greater intensity than the 532 bp band. This indicated that the 1.0 kb transcript was more abundant than the 1.2 kb transcript and was consistent with the northern blot analysis (Chapter 3). The ratio of the 602 bp band versus the 532 bp band appeared to remain constant, regardless of the relative level of *Hesx1* expression.

Figure 4.6 RNAase protection analysis of *Hesx1* expression in uninduced and induced ES cells.

RNAase protection analysis was carried out on 20 μg RNA using the Hesx1a derived antisense riboprobe spanning positions 26-628 of the Hesx1a cDNA sequence (see Fig. 4.5). The Hesx1a and Hesx1b bands correspond to the 1.0 kb and 1.2 kb *Hesx1* transcripts, respectively. RNA samples were as follows: ES-uninduced ES cells; SP-spontaneously induction: RA-Retinoic acid (RA) induction (low = 2 μM, high = 10 μM); MB-3-Methoxybenzamide (MBA) induction; DM-Dimethylsulfoxide (DMSO) induction; X+1-X cells +LIF, 1 passage; X-1-X cells -LIF, 1 passage, X+3-X cells + LIF, 3 passages. Uninduced and induced ES cell RNA was derived from MBL-5 ES cells (Pease *et al.*, 1990). Murine glyceraldehyde phosphate dehydrogenase (mGAP) (2.3.21) was used as a loading control. RNAase protection products were electrophoresed on a 5% polyacrylamide gel and autoradiographed for at -80°C 14 days with an intensifying screen.


low high
ES SP RA RA MB DM
Probe
Hesx1a
Hesx1b
mGAP

4.3.2 Analysis of *Hesx1* Expression in Different ES Cell Lines

Analysis of *Hesx1* expression described in sections 3.4.3. and 4.3.1 was undertaken exclusively using the MBL-5 ES cell line. RNAase protection analysis of *Hesx1* expression was carried out on a range of ES cell lines, all of which have been shown to be capable of generating germline chimeric mice. These ES cell lines included E14 (Handyside *et al.*, 1989), E14/TG2a (Hooper *et al.*, 1987), MBL-5 (Pease *et al.*, 1990), CCE (Schwartzberg *et al.*, 1989), J1 (Lyn Corchrine, Walter and Eliza Hall Institute, Melbourne, Australia), CGR8 (Mountford *et al.*, 1994), D3 (Doetschmann *et al.*, 1985) and CP1 (Bradley *et al.*, 1984, CP1 ES cell RNA was a kind gift from Dr. Austin Smith, CGR, Edinburgh). ES cell cultures used in this experiment contained greater than 99% undifferentiated stem cells.

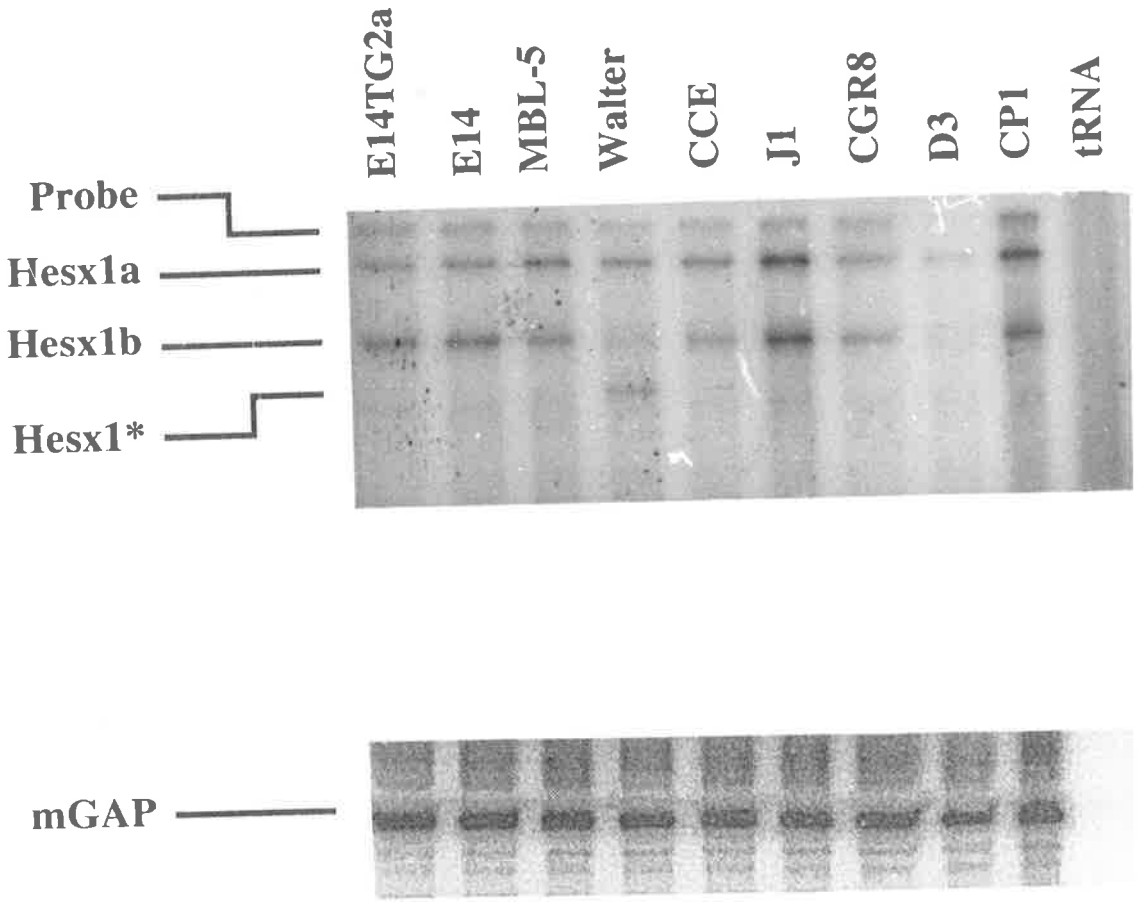
Hesx1 expression in the Walter ES cell line was also analysed. This ES cell line was of particular interest as it was isolated from a C57BL strain embryo in contrast to the other ES cell lines which were derived from 129/Sv strain mice. The Walter ES cell RNA was a kind gift from Dr. Austin Smith (CGR, Edinburgh).

RNAase protection analysis was carried out using 10 µg ES cell RNA. 602 bp (Hesx1a) and 532 bp (Hesx1b) bands corresponding to the 1.0 kb and 1.2 kb *Hesx1* transcripts were detected in all of the ES cell lines which were analysed (Fig. 4.7). This confirmed that *Hesx1* was not peculiar to the MBL-5 ES cell line, and supported the potential expression of *Hesx1* in the pluripotent cells of the early embryo. Minor variation in the level of *Hesx1* expression between the different ES cell lines was observed. This was particularly apparent in the D3 ES cell line which expressed a lower level of *Hesx1* than the other lines. As D3 ES cells were identical in appearance to the other ES cell lines, the low *Hesx1* expression level did not appear to be related to colony morphology. It is not clear whether the variable *Hesx1* expression levels reflected subtle variation in culture conditions, or were an intrinsic property of each ES cell line.

Apart from the 602 bp and the 532 bp bands, the Walter ES cell line also displayed an additional band of approximately 490 bp (Hesx1*; Fig. 4.7). The detection of three bands indicated that two distinct *Hesx1* alleles were present: a 129/Sv-like allele, which generated the 602 bp and the 532 bp bands (from the 1.0 kb and 1.2 kb transcripts respectively), and

Figure 4.7 RNAase protection analysis of *Hesx1* expression in a range of independent ES cell lines.

RNAase protection analysis was carried out on 20 µg RNA using the Hesx1a derived antisense riboprobe spanning positions 26-628 of the Hesx1a cDNA sequence (see Fig. 4.5). The Hesx1a and Hesx1b bands correspond to the 1.0 kb and 1.2 kb *Hesx1* transcripts, respectively. Hesx1* indicates the 490 bp RNAase protection product derived from the polymorphic *Hesx1* allele (detected in the Walter ES cell line). A tRNA control is also shown. Murine glyceraldehyde phosphate dehydrogenase (mGAP) (2.3.21) was used as a loading control. RNAase protection products were electrophoresed on a 5% polyacrylamide gel and autoradiographed at -80°C for 3 weeks with an intensifying screen.


E14TG2a
E14
MBL-5
Walter
CCE
J1
CGR8
D3
CP1
tRNA
Probe
Hesx1a
Hesx1b
Hesx1*
mGAP

an alternative allele (termed the Walter allele) which generated a single 490 bp band. Given that the Walter ES cell line expressed both *Hesx1* transcripts (as shown by the 602 bp and 532 bp bands), the detection of a single additional band (490 bp) suggested that the Walter allele had a sequence polymorphism immediately 5' of the point of divergence between the *Hesx1* transcripts. The difference in size between the Hesx1b (corresponding to the *Hesx1* transcript common region) and the Hesx1a* signals indicated that the sequence polymorphism was approximately 42 bp (532-490) upstream of the point of sequence divergence in the *Hesx1* transcripts. Unfortunately, insufficient Walter RNA was available to confirm this prediction. However, RNAase protection analysis of embryonic tissue RNA samples (derived from CBA strain embryos) also generated a 490 bp protection product which was characterized by sequence analysis (4.4).

4.3.3 Tissue Specific Analysis of *Hesx1* Expression

Hesx1 expression during embryogenesis was investigated by RNAase protection of tissue specific RNA samples. Unfortunately, facilities for the isolation of embryonic RNA from pluripotent cells during early development were not available. However, *Hesx1* expression at later stages of embryogenesis was investigated by RNAase protection analysis using tissue specific RNA samples isolated from 14.5 days p.c. (post-coitum) and 16 days p.c. embryos, and total embryonic RNA derived from 10.5 days p.c., 12.5 days p.c. and 16 days p.c. embryos (Fig. 4.8A,B). RNA samples from 10.5 days p.c., 12.5 days p.c., 16 days p.c. embryos and adults were isolated from CBA strain mice and were prepared with the help of Mr T. Schulz. 14.5 days embryonic tissue RNA samples were a kind gift from Dr. Austin Smith and were derived from MF1 (outbred albino) embryos.

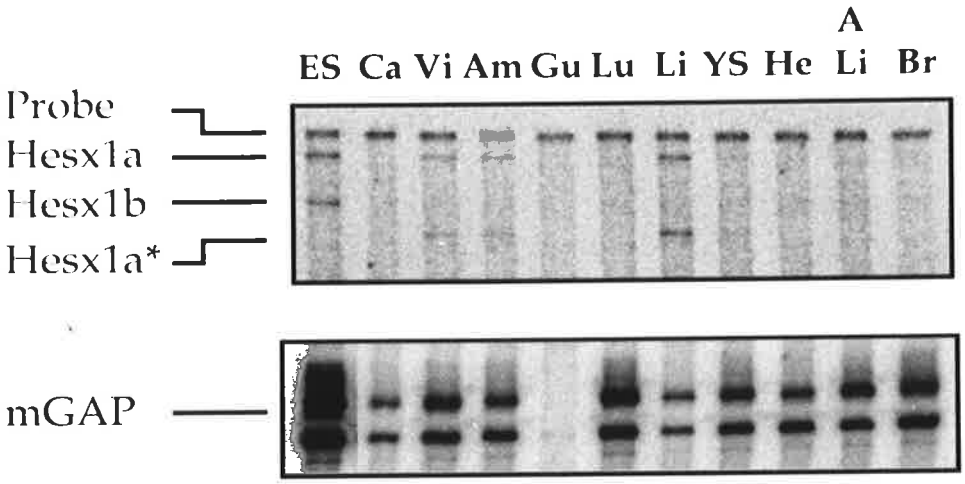
Hesx1 was not detected in total RNA isolated from the 10.5 days p.c. and 12.5 days p.c. embryos, and was expressed at low levels in the 16 days p.c. embryonic sample (Fig. 4.8B). Of the embryonic tissues examined in the 14.5 days p.c. and 16 days p.c. embryos, *Hesx1* expression was highest in the embryonic liver (Fig. 4.8A,B), and was detected at lower levels in the viscera (Fig. 4.8A). *Hesx1* expression in the embryonic liver was detected at similar levels to those seen in undifferentiated ES cells. It is not clear from this

Figure 4.8 RNAase protection analysis of *Hesx1* expression in embryonic and adult tissues.

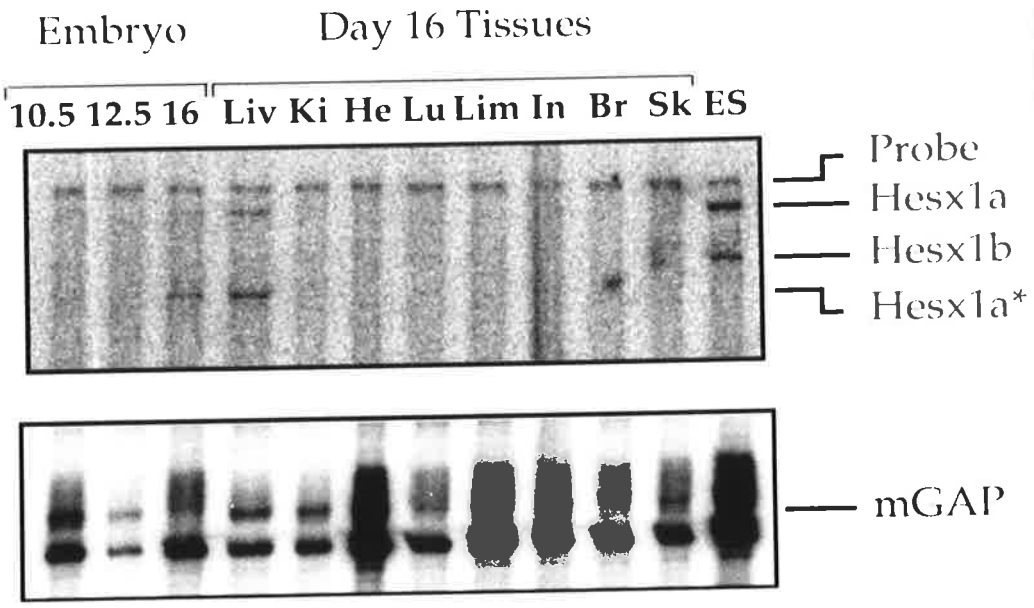
RNAase protections were carried out using the Hesx1a cDNA probe as described in Fig. 4.5. Hesx1a* indicates the Hesx1a transcript from the polymorphic allele. Murine glyceraldehyde phosphate dehydrogenase (mGAP) (2.3.21) was used as a loading control. As mGAP expression has been shown to vary considerably between different tissues (Robertson, *et al.*, 1993), and RNA samples were quantitated prior to RNAase protection, differences in mGAP signal between tissues does not reflect different amounts of RNA. RNAase protection products were separated on a 5% polyacrylamide gel and exposed to phosphorimage screens for 48 hours.

A. Embryonic samples were isolated from 14.5 days p.c. MF1 (outbred albino) embryos. Ca, calvaria; Vi, viscera; Am, amnion; Gu, gut; Lu, lung; Li, liver; YS, yolk sac; He, heart; Br, brain. ES, undifferentiated MBL-5 ES cells (Pease *et al.*, 1990). Each reaction contained 10 μg RNA. **B.** Total embryonic RNA was isolated from 10.5 days p.c., 12.5 days p.c. and 16 days p.c. CBA embryos. Embryonic samples from 16 days p.c. embryos are abbreviated as follows: Liv, liver; Ki, kidney; He, heart; Lu, lung; Lim, limbs; In, intestine; Br, brain; Sk, skin. ES, undifferentiated MBL-5 ES cells. Each protection reaction contained 20 μg RNA.

A


ES Ca Vi Am Gu Lu Li YS He A Li Br
Probe
Hesx1a
Hesx1b
Hesx1a*
mGAP

B


Embryo
Day 16 Tissues
10.5 12.5 16 Liv Ki He Lu Lim In Br Sk ES
Probe
Hesx1a
Hesx1b
Hesx1a*
mGAP

analysis whether this reflects ES-level expression in the majority of embryonic liver cells, or elevated expression levels in a sub-population of liver cells. *Hesx1* expression was also detected at low levels in the amnion and yolk sac which form part of the extra-embryonic tissue. *Hesx1* expression was not detected in adult liver (Fig. 4.8A) or in adult kidney, heart, lung, muscle or spleen (data not shown).

The detection of *Hesx1* expression in embryonic RNA samples derived from 14.5 days p.c. and 16 days p.c. embryos indicated that *Hesx1* is expressed during middle and late embryogenesis. Furthermore, restricted expression in the liver, viscera, amnion and yolk-sac suggested that *Hesx1* may play a developmental role in these tissues, and is consistent with the established spatial (and temporal) restriction of homeobox gene expression during embryogenesis (1.6).

In all embryonic tissues in which *Hesx1* was detected, the 532 bp band (corresponding to the 1.2 kb transcript) was replaced by a smaller protection product of approximately 490 bp. The absence of the 532 bp band indicated that the 1.2 kb *Hesx1* transcript, which was expressed by all of the ES cell lines (including Walter), was not expressed by the embyonic tissues. The different expression profile of the *Hesx1* transcripts in ES cells and in the embryo derived samples, suggested that these transcripts may be differentially regulated during embryogenesis. In particular, expression of the 1.2 kb *Hesx1* transcript may be restricted to the embryonic equivalent of ES cells, and may therefore carry out a distinct developmental function in these cells.

Given that the 1.2 kb *Hesx1* transcript was not detected in the embryonic RNA samples, the smaller protected product of approximately 490 bp (Hesx1a* Fig. 4.8) corresponded to a 1.0 kb transcript derived from an alternative *Hesx1* allele. This indicated that two distinct *Hesx1* alleles were present in the CBA and MF1 (outbred albino) mice which gave rise to the Hesx1a and Hesx1a* protection products. The upper protected species in the embryonic and the MBL-5 ES cell line samples were the same size, indicating that one of the CBA and MF1 (outbred albino) *Hesx1* alleles was identical or similar to the 129/Sv alleles (a 129/Sv-like allele). As the Hesx1a* protected bands in Figures 4.8 and 4.7 appeared to be identical in size, it is likely that this signal reflected a common *Hesx1*

polymorphism (i.e. a Walter allele) shared by the C57/BL, CBA and MF1 (outbred albino) strains.

4.4 Molecular Characterization of the *Hesx1* Polymorphism Identified by RNAase Protection in CBA Strain Mice

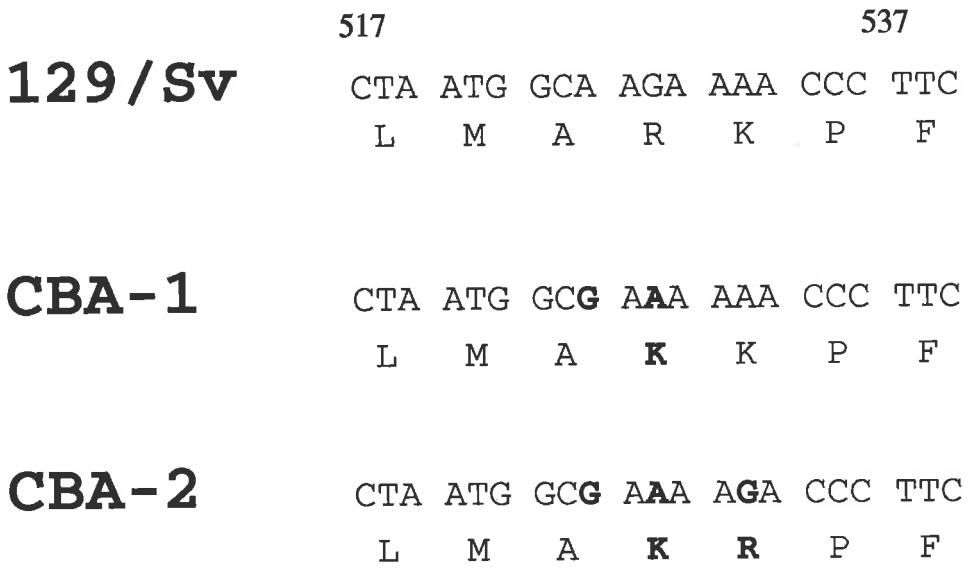
The molecular origin of the Hesx1a* RNAase protection product from the CBA embryonic RNA samples was investigated by PCR and sequence analysis. PCR was carried out on cDNA generated from days 16 p.c. embryonic liver RNA using primers which flanked the expected location of the polymorphic sequence (see 4.3.2). The upstream primer (which was identical to the *Wc* 1/5 primer used in the *Waved coat* analysis, see Chapter 5) was complimentary to the Hesx1a cDNA sequence spanning positions -54 to -41 (Fig. 4.2A). The downstream primer (3'Hesx1a*), which was specific for the 1.0 kb *Hesx1* transcript, was complimentary to positions 569 to 583 of the Hesx1a cDNA (Fig. 4.2A). The 653 bp PCR product generated from these primers was ligated into pT7T319U, and 2 independent clones were sequenced.

Comparison between the *Hesx1* cDNA clone sequences, derived from the 1.0 kb transcript, revealed that the clones were not identical. This implied the existence of two distinct *Hesx1* alleles (CBA-1 and CBA-2) which corresponded to the alternative cDNA sequences (Fig. 4.9). The identification of two distinct cDNA sequences derived from the 1.0 kb *Hesx1* transcript, was consistent with the detection of two RNAase protection products (Hesx1a and Hesx1a*) using the Hesx1a derived riboprobe.

The cDNA sequences were compared to the Hesx1a cDNA sequence (Fig. 4.9). The CBA-1 allele contained an A to G substitution and a G to A substitution at positions 525 and 527 respectively. The CBA-2 allele was identical to the CBA-1 allele apart from an additional A to G substitution at position 530. Interestingly, despite the generation of a full-length (602 bp) RNAase protection product from the 16 days p.c. liver RNA (Fig. 4.8B), neither the CBA-1 and CBA-2 alleles were identical to the 129/Sv *Hesx1* allele (Fig. 4.2A). Given that the CBA-1 allele had only two nucleotide differences from the 129/Sv allele, both of which were also present in the CBA-2 allele, it seems likely that the Hesx1a and the Hesx1a* RNAase protection products were derived from the CBA-1 and CBA-2 alleles,

Figure 4.9 Sequence comparison of the 129/Sv, CBA-1 and CBA-2 *Hesx1* alleles.

Nucleotide and predicted amino acid sequence of the 129/Sv, CBA-1 and CBA-2 *Hesx1* alleles, corresponding to positions 517 to 537 of the Hesx1a cDNA sequence. Nucleotide and amino acid residues of the CBA-1 and CBA-2 which differ from the 129/Sv sequence are shown in bold.


517
537
129/Sv
CTA ATG GCA AGA AAA CCC TTC
L M A R K P F
CBA-1
CTA ATG GCG AAA AAA CCC TTC
L M A K K P F
CBA-2
CTA ATG GCG AAA AGA CCC TTC
L M A K R P F

respectively. Presumably, the mismatches at positions 525 and 527 were not cleaved in the RNAase digestion reaction.

The nucleotide substitutions in the CBA-1 and CBA-2 alleles resulted in amino acid substitutions at one and two positions in the *Hesx1* open reading frame, respectively (Fig. 4.9). The G to A substitution at position 527 (which was common to both alleles) resulted in a substitution of a lysine for an arginine at position 176 of the open reading frame. The unique CBA-2 A to G substitution at position 530 induced a lysine to arginine substitution at position 177 of the open reading frame. Therefore, the 129/Sv, CBA-1 and CBA-2 *Hesx1* alleles encode highly related but distinct open reading frames. As the sequence polymorphisms identified in the CBA-1 and CBA-2 *Hesx1* alleles resulted in conservative amino acid substitutions, it is possible that the charge of amino acid residues at these positions is important for Hesx1 structure and/or function.

It should be noted that only single clones corresponding to the CBA-1 and CBA-2 alleles were characterized. This raises the possibility that some or all of the sequence variations which were detected, may be due to PCR error. This possibility seems unlikely as the location of the base substitutions was consistent with the estimated length of the Hesx1a* band. However, this possibility of PCR induced error could be excluded by characterization of additional clones.

4.5 Discussion

Screening 1.25 x 10^6 plaques from a D3 ES cell cDNA library yielded a total of five *Hesx1* cDNA clones. Four of these clones, which contained 847 bp of *Hesx1* sequence (Hesx1a), corresponded to the more abundant 1.0 kb transcript. The remaining clone (Hesx1b), which was derived from the 1.2 kb (Hesx1b) transcript, contained 1117 bp of *Hesx1* sequence. The smaller size of the *Hesx1* cDNA clones, compared with their corresponding transcripts, indicated the cDNA clones were not full-length. Given that the Hesx1a clone had a $(A)_n$ tail, additional sequence corresponding to the 1.0 kb *Hesx1* transcript must be located at the 5' end of this clone. The Hesx1b clone did not contain a $(A)_n$ tail, and therefore additional 1.2 kb *Hesx1* transcript sequence may be located at either

end of this clone. Given that the Hesx1a and Hesx1b clones differ in length at the 3' end, it seems likely that the 1.0 kb and 1.2 kb *Hesx1* transcripts are 5' coterminal. The isolation of further 5' and 3' cDNA sequence, by rescreening the ES cell cDNA library or by RACE PCR (Frohman *et al.*, 1988), would be required to confirm this. Additional 5' Hesx1a cDNA sequence would also confirm whether the upstream in-frame stop codon spanning positions -72 to -70 of the Hesx1b cDNA clone (1.2 kb transcript) is also encoded by the 1.0 kb *Hesx1* transcript.

A $(CA)_n$ repeat sequence, containing 12 direct dinucleotide repeat units, was identified in the 3' untranslated region of the Hesx1b cDNA clone. Although these short repeat sequences occur at high frequency within the murine genome (1×10^5 copies/haploid genome; Hamada and Kakunaga, 1982), their genomic distribution appears to be random, and their biological function remains obscure (Stallings *et al.*, 1991). The high frequency and random genomic distribution of $(CA)_n$ repeats has been exploited to generate a genetic linkage map of the mouse genome based on polymorphic variation in the number of repeat units (Dietrich *et al.*, 1992). Given that the *Hesx1* CA contains 12 dinucleotide repeats, and is therefore likely to be polymorphic (Weber, 1990), the *Hesx1* repeat may provide an additional polymorphic marker for this type of genomic mapping.

Comparison between the *Hesx1* and partial XANF-1 ORFs revealed 47% homology across the entire *Hesx1* ORF, including 80% identity within the homeodomain. This degree of homology is probably not sufficient to class *Hesx1* and XANF-1 as sequence homologs, but it is likely, given that these genes also share structural similarities, that *Hesx1* and XANF-1 are derived from a common ancestral gene. It was interesting that homology between *Hesx1* and XANF-1 was not restricted to the homeodomain, but was also detected in the homeodomain flanking sequence and, to a lesser extent, at the amino terminus. The conservation of these additional regions of homology suggests that these sequences may play a role in protein structure and/or function.

The sequence and structural conservation between *Hesx1* and XANF-1 may reflect conserved developmental roles for these related genes during murine and *Xenopus* embryogenesis. Murine and *Xenopus* homologs of the *goosecoid* (Blum *et al.*, 1992; Cho

et al., 1991) and *Nkx2.5* (Lints *et al.*, 1993; Tonissen *et al.*, 1994) homeobox genes have been isolated. These homeobox gene homologs share 78% and 62% identity, respectively. Evidence based on gene expression and function suggests that the *goosecoid* and *Nkx-2.5* homologs fulfil similar roles in these different species during development. Further investigation into the expression and function of *Hesx1* and XANF-1 during murine and *Xenopus* development is therefore required to determine whether these genes also represent functional homologs.

In addition to the homeodomain, homeobox genes may also encode other conserved structural motifs including the POU domain (Scholer, 1991), zinc-fingers (Fortini *et al.*, 1991) and the paired box (Gruss and Walther, 1992). However, analysis of the *Hesx1* open reading frame failed to detect any sequences corresponding to known conserved structural motifs (besides the homeodomain). Short conserved amino acid sequences have also been detected in the genes which belong to some homeobox classes, in particular the *Antennapedia* class (Burglin, 1994). Given that *Hesx1* and XANF-1 share regionalized homology outside the homeodomain sequence, it is possible that the isolation of further homeobox genes which belong to this class may allow identification of additional conserved sequences.

Analysis of *Hesx1* expression in ES cells and induced ES cells was investigated using RNAase protection. *Hesx1* was expressed at lower levels in the chemically- and spontaneously-induced ES cells, which indicated that *Hesx1* was downregulated during ES cell differentiation. This was consistent with the northern blot analysis of *Hesx1* expression (Chapter 3) and supported a possible role for *Hesx1* in the regulation of pluripotent cell differentiation and renewal during early embryogenesis.

RNAase protection analysis indicated that the 1.0 kb and 1.2 kb *Hesx1* transcripts were expressed in all of the ES cells lines which were assayed, albeit with minor variation in expression levels between different lines. The detection of *Hesx1* expression in multiple independent ES cell lines showed that *Hesx1* expression was not restricted to MBL-5 ES cells, and supported a developmental role for *Hesx1* in the pluripotent cells of the early mouse embryo. *Hesx1* expression in the D3 ES cells was lower than the other ES cell lines.

Although the reason for this reduced *Hesx1* expression is not known, RNAase protection analysis of these ES cell RNA samples has indicated that *MMoxB* expression (which also appears to be restricted to undifferentiated ES cells, Chapter 3) was also reduced in the D3 ES cell RNA sample (*MMoxB* expression in the other lines was higher and relatively constant; Gavin Chapman, University of Adelaide, personal communication). Further analysis of *Hesx1* expression in additional cultures of D3 ES cells is therefore required to determine whether the low *Hesx1* expression levels were an intrinsic property of this cell line, or were simply an artifact of this particular D3 ES cell sample.

The possibility that *Hesx1* may play a developmental role during post-gastrulation embryogenesis, was investigated by RNAase protection analysis of *Hesx1* expression in tissue specific RNA samples derived from 14.5 days p.c. and 16 days p.c. embryos. Analysis of *Hesx1* levels in embryonic tissues indicated that that *Hesx1* was expressed during 14.5 days p.c. and 16 days p.c. of embryogenesis and at highest levels in the embryonic liver. During embryogenesis the liver first becomes apparent around 11.5-12 days p.c., and replaces the yolk-sac as the principal site of haematopoiesis with concomitant establishment of the definitive red blood cell lineage (Kaufman, 1992). The haematopoietic function of the liver continues until about 18 days p.c., when the spleen and subsequently the bone marrow take over this role. The restricted *Hesx1* expression in embryonic liver, coupled with the absence of *Hesx1* expression in adult mouse liver which does not have an haematopoietic function, suggest that *Hesx1* might be expressed by cells of the haematopoietic lineages during embryogenesis. The levels of *Hesx1* expression within the days 16.5 p.c. embryonic liver were similar to the levels expressed by undifferentiated ES cells *in vitro*. This could reflect generalised expression in all liver cells, or elevated *Hesx1* expression levels in a sub-population of embryonic liver cells such as specific haematopoietic lineages. It will therefore be of interest to establish the expression pattern and function of *Hesx1* during haematopoiesis. The lower levels of *Hesx1* expression detected in the amnion and yolk-sac may be due to the presence of haematopoietic stem cells which are believed to migrate from the wall of the yolk-sac to the liver to establish haematopoietic foci (Kaufman, 1992). *Hesx1* expression was also detected at low levels in

the viscera which is comprised of the embryonic tissues which remain after removal of the major organs. The detection of *Hesx1* expression in the viscera may be due to the presence of haematopoietic cells within these remaining tissues or may indicate additional sites of *Hesx1* expression in the embryo.

The 1.2 kb *Hesx1* transcript, which was expressed by ES cells, was not detected in the tissue specific RNA samples derived from 14.5 days p.c. and 16 days p.c. embryos. This suggested that expression of the 1.2 kb transcript may be restricted to pluripotent cells during early development. This could be investigated by *in situ* analysis of *Hesx1* expression in the early embryo, using transcript specific probes. Given that the 1.0 kb and 1.2 kb transcripts encode identical open reading frames, the possible unique biological role of the 1.2 kb transcript remains obscure.

The detection of a 490 bp *Hesx1* RNAase protection product (Hesx1*) in RNA derived from C57/BL (Walter ES cells), MF1 (outbred albino; 14.5 days p.c. embryonic RNA) and CBA (16 days p.c. embryonic RNA) strain mice, indicated that the *Hesx1* locus was polymorphic. Sequence analysis of the of CBA strain identified two distinct *Hesx1* alleles, CBA-1 and CBA-2. Given that the Hesx1a* band derived from the C57/BL (Walter) *Hesx1* allele and the CBA-2 *Hesx1* allele were of similar size, it seems likely that these alleles were identical. The CBA-1 and CBA-2 alleles had one and two conservative substitutions of basic amino acids at positions 176 and 177, respectively. As these residues are not located within the homeodomain, their possible role is not not known. However, it seems likely that the charge conservation at these positions, reflects a functional role for these residues in the *Hesx1* protein.

It was interesting that the 1.0 kb *Hesx1* transcript, derived from the CBA-1 allele, differed in sequence at two positions from the (129/Sv allele derived) riboprobe. The inability of the RNAase protection reaction to discriminate between the 129/Sv and CBA-1 derived 1.0 kb *Hesx1* transcripts, raises the possibility that the 602 bp RNAase protection signal detected in the MF1 (outbred albino) and C57/BL RNA samples may also be derived from a CBA-1 *Hesx1* allele. Alternatively, this signal may correspond to novel *Hesx1*

allele. This possibility may be investigated by PCR and sequence analysis as has been carried out for the CBA strain.

CHAPTER 5: CHARACTERIZATION OF THE *Hesx1* LOCUS

CHAPTER 5: CHARACTERIZATION OF THE *Hesx1* LOCUS

5.1 Introduction

As described in Chapter 4, *Hesx1* was selected for further analysis on the basis of downregulated expression during ES cell differentiation and novel sequence. Attempts to isolate a *Hesx1* genomic clone were carried out concurrently with the isolation of the *Hesx1* cDNA clones (Chapter 4). The isolation of a genomic clone was desirable as this would allow characterization of the *Hesx1* locus and, in particular, allow identification of the molecular origins of the alternative *Hesx1* transcripts.

This chapter also describes the physical mapping of the *Hesx1* locus. Mapping of homeobox gene loci has led to the identification of the molecular basis for at least two murine developmental mutations (Gruss and Walther, 1992). The Sp^{2H} allele of the murine developmental mutation *splotch* (*Sp*), has a 32 nucleotide deletion within the *Pax-3* homeobox, resulting in premature termination of the open reading frame via a newly created stop codon at the deletion breakpoint (Epstein *et al.*, 1991). Mice homozygous for the Sp^{2H} allele die before birth due to abnormalities in many of the neural tissues in which *Pax-3* is expressed during development (Epstein *et al.*, 1991), including the neural tube and certain neural crest cell derivatives (Goulding *et al.*, 1991). Two alleles of the *Small eye* (*Sey*) developmental mutation, *Sey* and Sey^{Neu}, have been shown to carry mutations in the homeobox gene *Pax-6*, which result in premature termination of the open reading frame (Hill *et al.*, 1991). As the *Pax-6* gene is expressed by the eye during development (Gruss and Walther, 1992; Hill *et al.*, 1991), the absence of eye structures in *Sey* homozygous mice presumably results from an inability of the truncated *Pax-6* protein to carry out its developmental function (Hill *et al.*, 1991). Identification of developmental mutation(s) which co-localize with the *Hesx1* locus, and the subsequent molecular analysis of *Hesx1* in these mutant strain(s), could therefore provide insight into the role of *Hesx1* during development. Mapping of the *Hesx1* locus would also determine whether *Hesx1* is located within any of the four *Hox* clusters on chromosomes 2, 6, 11 and 15. This was considered

to be unlikely as *Hesx1* has little homology with the *Antennapedia* class homeobox genes which constitute the *Hox* clusters (Chapter 4).

Genomic southern blot analysis and isolation of the *Hesx1* genomic clone were carried out in conjunction with Mr. B Johnson (Johnson, 1992). The chromosomal localization of *Hesx1* was performed by Dr. G Webb (Department of Genetics, Queen Elizabeth Hospital).

5.2 Isolation and Characterisation of a 16 kb *Hesx1* Genomic Clone

5.2.1 Restriction Map of the *Hesx1* Locus

Prior to isolation of a *Hesx1* genomic clone, the *Hesx1* genomic region was characterized southern blot analysis. Ten μg murine genomic DNA isolated from MBL-5 embryonic stem (ES) cells (Pease *et al.*, 1990; strain 129/Sv) was digested with *Bam* HI, *Hind* III, *Pst* I and *Eco* RI in both single and double digests, electrophoresed on a 1% agarose gel and blotted. The genomic southern was probed with the 147 bp *Hesx1* PCR fragment containing the partial homeobox sequence and flanking primers (3.2.1) and washed in 1 X SSC, 0.1% SDS at 42°C for 45 minutes (Fig. 5.1A). Single bands detected in each track were used to deduce a restriction map of the *Hesx1* locus (Fig. 5.1B). The absence of multiple bands indicated that homeobox genes closely related to *Hesx1* were not present in the murine genome. *Hesx1* may therefore represent the only murine member of the novel homeodomain sequence class defined by *Hesx1* and XANF-1 (Chapter 4).

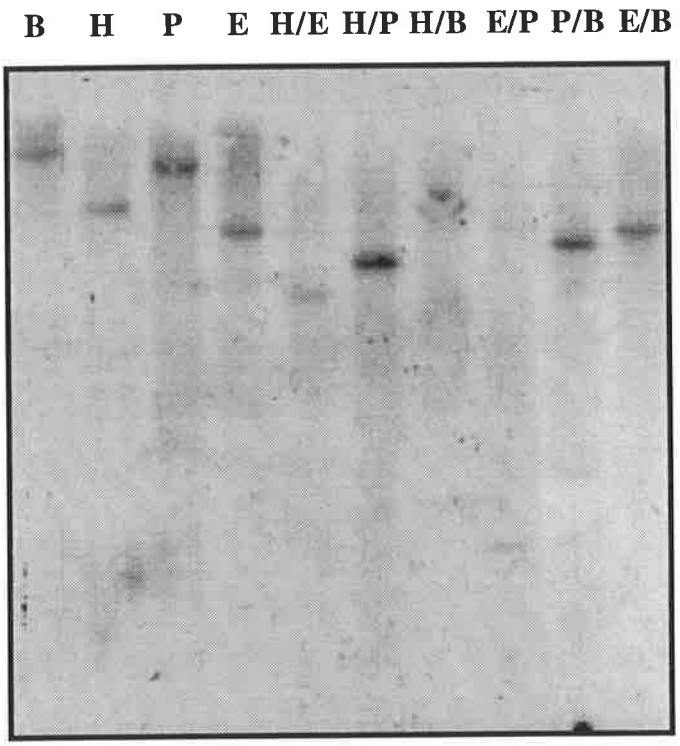
5.2.2 Isolation of a *Hesx1* Genomic Clone

To isolate a *Hesx1* genomic clone, a murine genomic library derived from BALB/c strain mice was screened using the *Hesx1* partial homeobox sequence (as for 5.2.1). 2.5 X 10^5 plaques were screened and a single third round positive plaque was isolated and purified. This phage, named Hesx1g, contained a 16 kb insert spanning the *Hesx1* locus. The restriction map for the 16 kb insert was determined from southern blot analysis of Hesx1g digested with *Bam* HI, *Hind* III, *Pst* I and *Eco* RI probed with the *Hesx1* partial

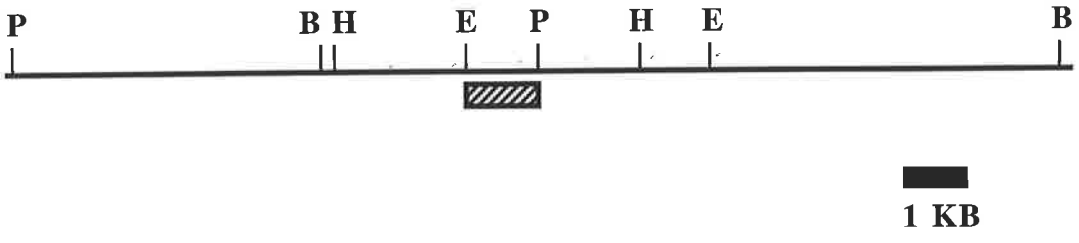
Figure 5.1 Genomic southern blot of the *Hesx1* locus.

A. 10 μg murine genomic DNA was digested, electrophoresed on a 1% agarose gel and blotted. The southern blot was probed with oligolabelled *Hesx1* PCR product generated by the En5 and En3 primers (3.2.1) and washed in 1 X SSC / 0.1% SDS at 42°C for 45 minutes. The approximate sizes of the bands are: B- 12 kb; H- 5 kb; P- 9 kb; E- 4 kb; H/E- 2.9 kb; H/P- 3.3 kb; H/B- 5 kb; E/P- 1.2 kb, P/B- 3.6 kb; E/B- 4 kb. Note that the 5 kb H/B band is obscured and that the 1.2 kb E/P band was visible on the original autoradiograph. E=*Eco* RI; B=*Bam* HI; H=*Hind* III; P=*Pst* I. **B.** *Hesx1* genomic restriction map deduced from the genomic southern. The genomic region which contains the homeobox is shown by the striped box.

A



B



homeobox sequence (data not shown). This restriction map was identical to the *Hesx1* genomic map (Fig. 5.1B) which confirmed the authenticity of the genomic clone (Fig. 5.2).

5.2.3 Identification of *Hesx1* Intron/Exon Boundaries

Hesx1 exons within the 16 kb genomic region were identified by hybridization of the Hesx1a and Hesx1b cDNAs to a southern blot carrying restriction digested Hesx1g (Johnson, 1992). Genomic regions which hybridized to the cDNA probes were subcloned into pT7T319U vector, which generated a series of clones which encompassed the *Hesx1* genomic region (Johnson, 1992). Genomic clone regions from which *Hesx1* intron and exon sequence were generated are shown in Figure 5.2A. As the aim of this approach was to identify the position of the *Hesx1* intron/exon boundaries, the genomic region was not sequenced in its entirety.

Intron/exon boundaries were identified by comparison of the *Hesx1* cDNA and genomic sequence. Three introns (i, ii and iii) common to the Hesx1a and Hesx1b cDNAs were identified (Fig. 5.2A,B). These introns were flanked by conserved splice donor and splice acceptor sequences (Fig 5.3; Fig. 4.2) as defined by Mount (Mount, 1982). Introns ii and iii were located within the homeobox itself. Although the presence of a single intron within a homeobox is not uncommon, few examples of two introns in this region have been reported, and *Hesx1* is the first mammalian homeobox gene for which this has been described (Burglin, 1994).

The Hesx1a and Hesx1b cDNA clones (4.2.1) were smaller than the 1.0 kb and 1.2 kb *Hesx1* transcripts from which they were derived, and were therefore not full-length clones. As the Hesx1a cDNA clone had a $(A)_n$ tail, 5' sequence derived from the 1.0 kb *Hesx1* transcript was not represented in this clone. Although the Hesx1b cDNA clone did not have a $(A)_n$ tail, this clone may also be incomplete at the 5' end. It is possible, therefore, that additional upstream introns are present at the *Hesx1* locus which were not identified due to this lack of 5' cDNA sequence. The isolation of additional 5' *Hesx1* cDNA sequence is required to determine if this is the case.

Figure. 5.2 *Hesx1* genomic restriction map and intron/exon structure.

A. The restriction map was deduced from southern blot analysis of the 16 kb *Hesx1* genomic clone. Genomic regions which contained exons were subcloned and sequenced, and are indicated by arrows. Boxed regions represent *Hesx1* exons. The 5' boundary of exon I, and the 3' boundary of exon V, have not been identified and are represented by dashed lines. The length of each intron is indicated in italics in parentheses. Light shading represents the *Hesx1* open reading frame and darker shading the homeobox region. The alternative splice site, which is located within exon IV, is indicated by an asterisk. The exact positions of intron/exon boundaries are shown in Fig. 4.2. The 9 kb *Pst* I fragment used in the *Wc* southern blot analysis (5.4.2) is indicated by the striped box. H=*Hind* III, E=*Eco* RI, P=*Pst* I. **B.** Derivation of the 1.0 kb (Hesx1a) and 1.2 kb (Hesx1b) transcripts. The 1.2 kb transcript results from splicing between the weak splice doner site (asterisk) in exon IV to the alternative downstream exon V. The 1.0 kb transcript terminates within exon IV.

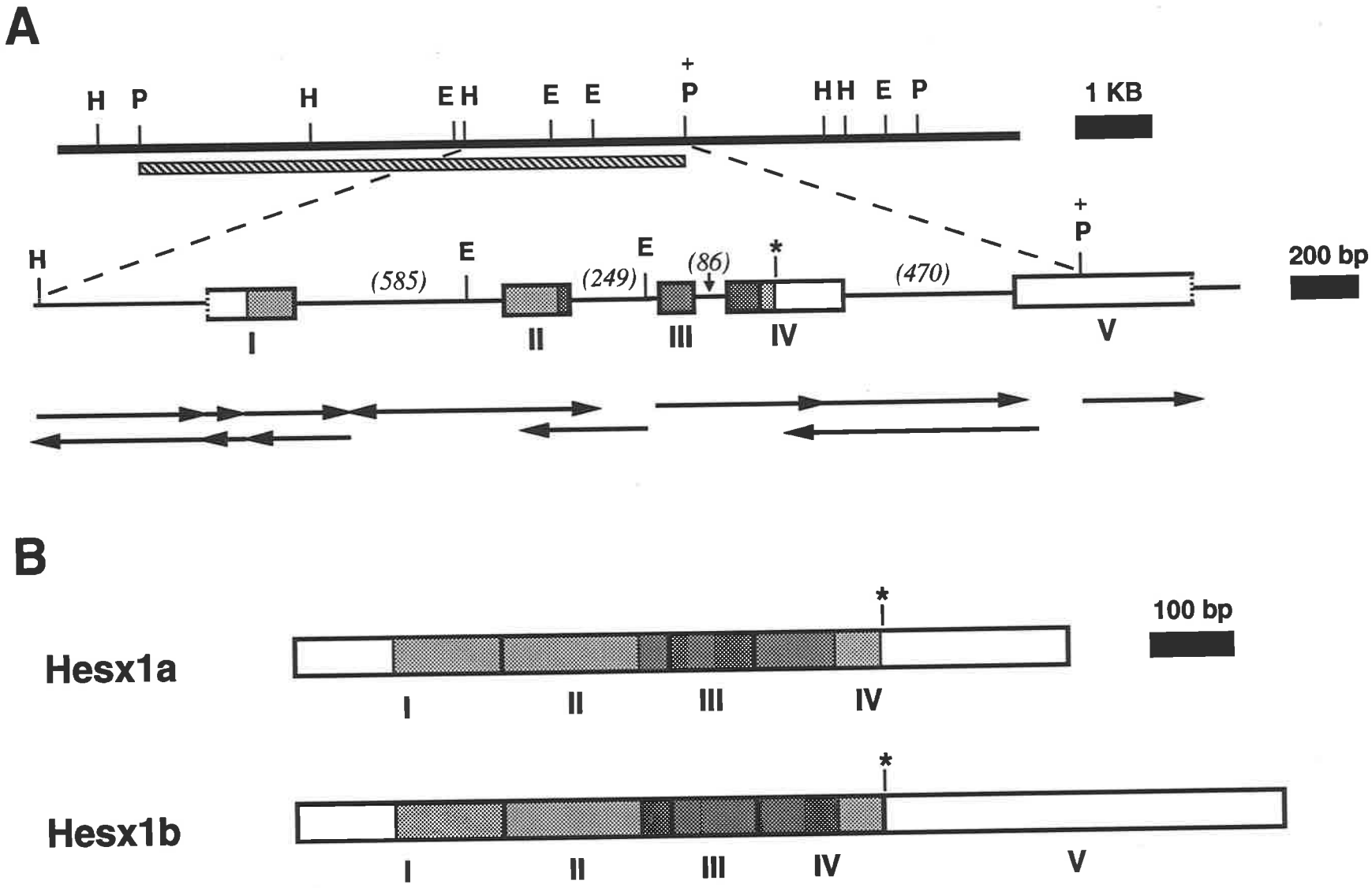

A
H P H E H E E + P H H E P
1 KB
H E E + P
(585) (249) (86) (470)
200 bp
I II III IV V
B
Hesx1a
Hesx1b
100 bp
I II III IV
I II III IV V

Figure 5.3 Compilation of the splice doner and splice acceptor sites for the four *Hesx1* introns.

4* refers to the doner splice site utilized by the Hesx1b transcript within exon 4. The length of intronic sequences (in base pairs) in shown in parentheses. The consensus splice doner and acceptor sites as defined by Mount (Mount, 1982) are also shown.

EXON	INTRON			EXON
N°	*5' SPLICE SITE*		*3' SPLICE SITE*	N°
1	CAG	GTATGC--- **1 (585)** ---TTCTTCTGATTTTAG	A	2
2	CAG	GTGAGA--- **2 (249)** ---TCCTTCTTATTCTAG	G	3
3	CAG	GTAGGA--- **3 (86)** ---TCCCCTTGATTCTAG	A	4
4*	TAG	GTAGAA--- *** (663)** ---ATTCGTTTTTTCCAG	G	5
	$\begin{matrix}C\\A\end{matrix}$AG	GT$\begin{matrix}A\\G\end{matrix}$AGT **CONSENSUS** $(\begin{matrix}T\\C\end{matrix})_{11}$ N$\begin{matrix}C\\T\end{matrix}$AG	G	

5.2.4 Molecular Origin of the 1.0 kb and 1.2 kb *Hesx1* Transcripts

Genomic regions specific to the Hesx1a and Hesx1b cDNA clones were identified by southern blot analysis. Hybridization of Hesx1a and Hesx1b cDNA probes to a southern blot of restriction digested Hesx1g, showed that only the Hesx1b probe hybridized to genomic sequences downstream of the *Pst* I site downstream of the homeobox (indicated by a cross in Fig. 5.2A; Johnson, 1992). This indicated that the Hesx1a 3' unique sequence was upstream of this *Pst* I site. Comparison of the genomic and Hesx1a cDNA sequences confirmed that the unique 3' end of the Hesx1a cDNA was contiguous with the *Hesx1* open reading frame and was contained entirely within exon IV (Fig 5.2A,B).

The genomic region corresponding to the 3' unique Hesx1b sequence (exon V) was identified 470 bp downstream of the 3' end of exon IV (Fig. 5.2A). A consensus splice acceptor site spanning the intron/exon boundary at the 5' end of exon V was identified (Fig. 5.3). Although the complete exon V sequence was not determined, southern blot analysis of the genomic clone indicated that the entire 3' end of the Hesx1b cDNA was contained within this exon (data not shown). As the Hesx1b cDNA was incomplete at the 3' end, it is likely that exon V contained additional 3' sequence which was not identified.

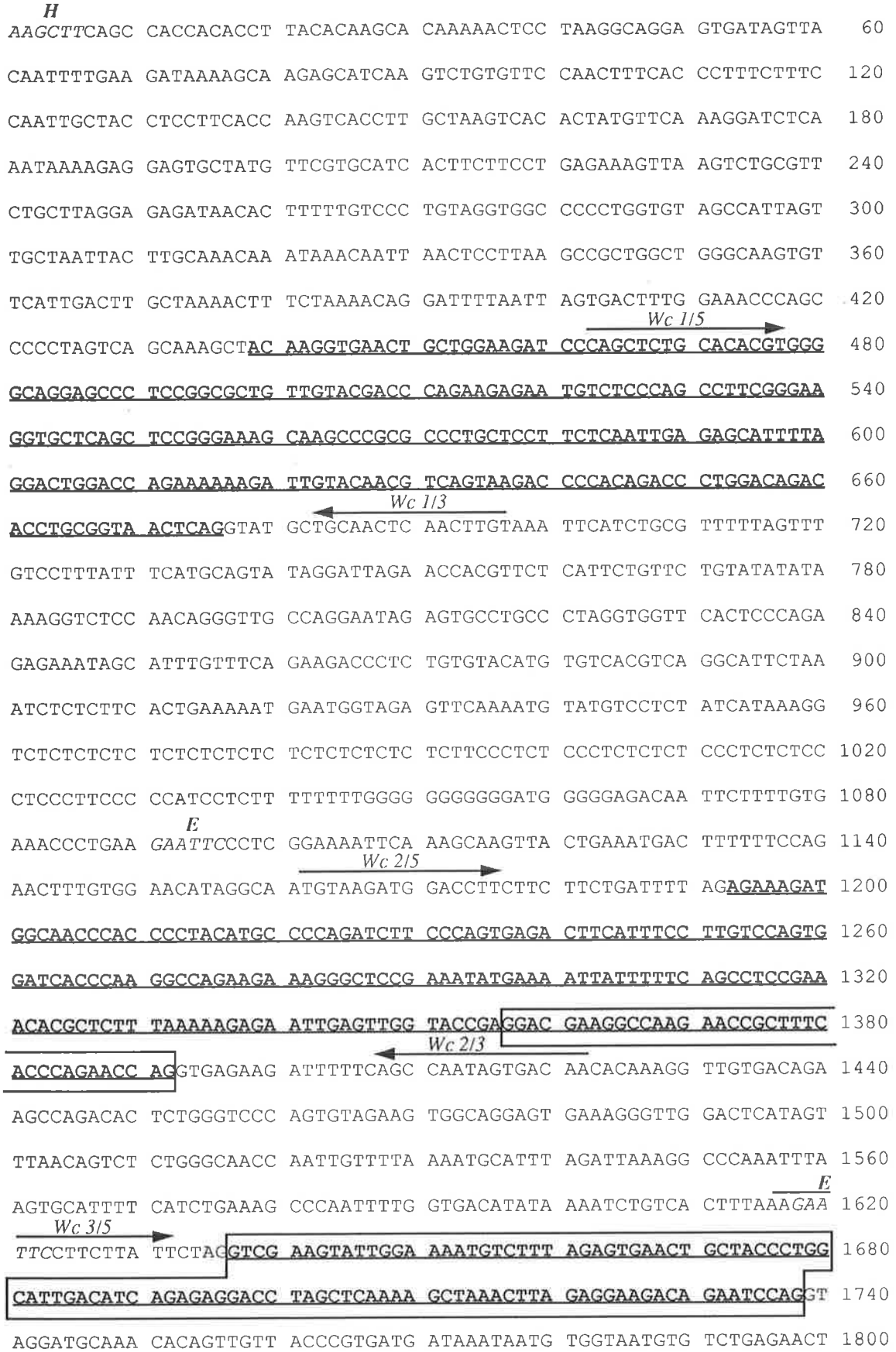
A poor consensus donor splice site, which was conserved at only 5 of the 9 positions (Fig. 5.3) was identified at the point of Hesx1a/Hesx1b divergence within exon IV. Therefore, the 1.2 kb *Hesx1* transcript was produced by utilization of this weak donor splice site to attach the downstream exon V (Fig. 5.2B; Fig. 4.2). As expression of the 1.2 kb transcript has been detected only in ES cells (Chapter 4), this alternative splicing event may be restricted to pluripotent cells. It is also possible that the poor consensus of this splice doner site contributes to the lower expression levels of the 1.2 kb *Hesx1* transcript compared with the 1.0 kb transcript in ES cells

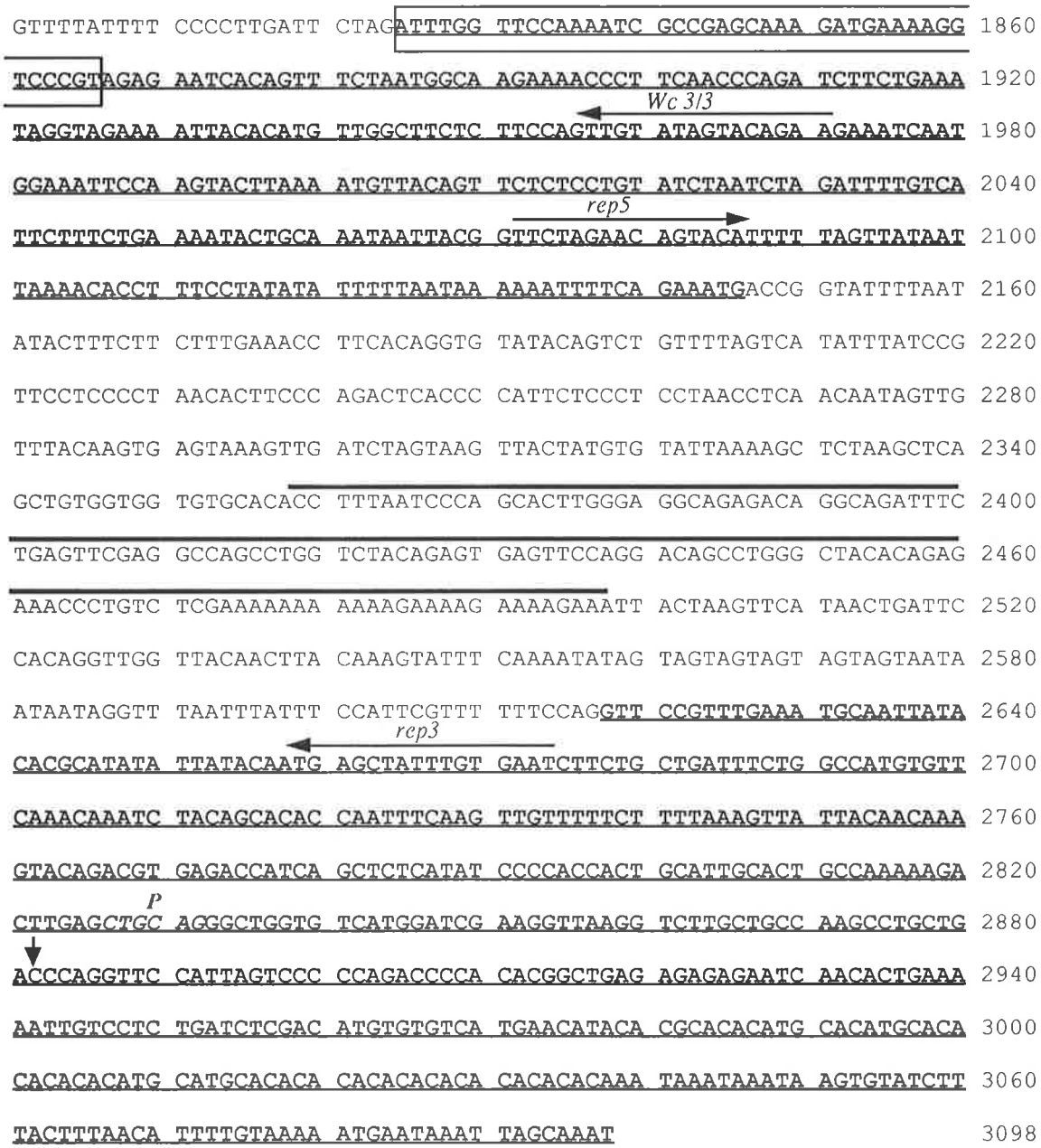
5.2.5 Comparison of *Hesx1* Genomic and cDNA Sequences

Apart from 168 bp within Exon V spanning positions 2658 to 2826 of the *Hesx1* genomic sequence, the *Hesx1* exons within the genomic clone were sequenced in their entirety (Fig. 5.4) and compared to the *Hesx1* cDNA sequences. The only difference

Figure 5.4 Nucleotide sequence of the *Hesx1* locus.

The genomic sequence spanning the *Hesx1* transcription unit was derived from the Hesx1g genomic clone (see Fig. 5.2A for sequenced regions). Exon sequences are bold and underlined and the homeobox is boxed. The T to C substitution within exon V is indicated by the arrow. The position of the PCR primers used in the *Waved coat* analysis (see 5.4.1) and the rep5 and rep3 sequencing primers (see 5.5) are indicated by the horizontal arrows. The B1 repeat sequence is shown under the bold overline. Restriction sites are shown in italics. *E=Eco* RI; *H=Hind* III; *P=Pst* I. Note that the Exon V sequence spanning positions 2658 to 2826 was derived from the Hesx1b cDNA clone.


H
AAGCTTCAGC CACCACACCT TACACAAGCA CAAAAACTCC TAAGGCAGGA GTGATAGTTA 60
CAATTTTGAA GATAAAAGCA AGAGCATCAA GTCTGTGTTC CAACTTTCAC CCTTTCTTTC 120
CAATTGCTAC CTCCTTCACC AAGTCACCTT GCTAAGTCAC ACTATGTTCA AAGGATCTCA 180
AATAAAAGAG GAGTGCTATG TTCGTGCATC ACTTCTTCCT GAGAAAGTTA AGTCTGCGTT 240
CTGCTTAGGA GAGATAACAC TTTTTGTCCC TGTAGGTGGC CCCCTGGTGT AGCCATTAGT 300
TGCTAATTAC TTGCAAACAA ATAAACAATT AACTCCTTAA GCCGCTGGCT GGGCAAGTGT 360
TCATTGACTT GCTAAAACTT TCTAAAACAG GATTTTAATT AGTGACTTTG GAAACCCAGC 420
Wc 1/5
CCCCTAGTCA GCAAAGCTAC AAGGTGAACT GCTGGAAGAT CCCAGCTCTG CACACGTGGG 480
GCAGGAGCCC TCCGGCGCTG TTGTACGACC CAGAAGAGAA TGTCTCCCAG CCTTCGGGAA 540
GGTGCTCAGC TCCGGGAAAG CAAGCCCGCG CCCTGCTCCT TCTCAATTGA GAGCATTTTA 600
GGACTGGACC AGAAAAAAGA TTGTACAACG TCAGTAAGAC CCCACAGACC CTGGACAGAC 660
Wc 1/3
ACCTGCGGTA ACTCAGGTAT GCTGCAACTC AACTTGTAAA TTCATCTGCG TTTTTAGTTT 720
GTCCTTTATT TCATGCAGTA TAGGATTAGA ACCACGTTCT CATTCTGTTC TGTATATATA 780
AAAGGTCTCC AACAGGGTTG CCAGGAATAG AGTGCCTGCC CTAGGTGGTT CACTCCCAGA 840
GAGAAATAGC ATTTGTTTCA GAAGACCCTC TGTGTACATG TGTCACGTCA GGCATTCTAA 900
ATCTCTCTTC ACTGAAAAAT GAATGGTAGA GTTCAAAATG TATGTCCTCT ATCATAAAGG 960
TCTCTCTCTC TCTCTCTCTC TCTCTCTCTC TCTTCCCTCT CCCTCTCTCT CCCTCTCTCC 1020
CTCCCTTCCC CCATCCTCTT TTTTTTGGGG GGGGGGGATG GGGGAGACAA TTCTTTTGTG 1080
E
AAACCCTGAA GAATTCCCTC GGAAAATTCA AAGCAAGTTA CTGAAATGAC TTTTTTCCAG 1140
Wc 2/5
AACTTTGTGG AACATAGGCA ATGTAAGATG GACCTTCTTC TTCTGATTTT AGAGAAAGAT 1200
GGCAACCCAC CCCTACATGC CCCAGATCTT CCCAGTGAGA CTTCATTTCC TTGTCCAGTG 1260
GATCACCCAA GGCCAGAAGA AAGGGCTCCG AAATATGAAA ATTATTTTTC AGCCTCCGAA 1320
ACACGCTCTT TAAAAAGAGA ATTGAGTTGG TACCGAGGAC GAAGGCCAAG AACCGCTTTC 1380
Wc 2/3
ACCCAGAACC AGGTGAGAAG ATTTTTCAGC CAATAGTGAC AACACAAAGG TTGTGACAGA 1440
AGCCAGACAC TCTGGGTCCC AGTGTAGAAG TGGCAGGAGT GAAAGGGTTG GACTCATAGT 1500
TTAACAGTCT CTGGGCAACC AATTGTTTTA AAATGCATTT AGATTAAAGG CCCAAATTTA 1560
E
AGTGCATTTT CATCTGAAAG CCCAATTTTG GTGACATATA AAATCTGTCA CTTTAAAGAA 1620
Wc 3/5
TTCCTTCTTA TTCTAGGTCG AAGTATTGGA AAATGTCTTT AGAGTGAACT GCTACCCTGG 1680
CATTGACATC AGAGAGGACC TAGCTCAAAA GCTAAACTTA GAGGAAGACA GAATCCAGGT 1740
AGGATGCAAA CACAGTTGTT ACCCGTGATG ATAAATAATG TGGTAATGTG TCTGAGAACT 1800


Wc 3/3
rep5
rep3
P

between the *Hesx1* cDNA and genomic clones was a T to C substitution at position 823 of the Hesx1b sequence (position 2882 of the genomic sequence) within Exon V. As the cDNA and genomic clones were isolated from strain 129/Sv and BALB/c libraries respectively, it seems likely that this single base substitution represents a sequence polymorphism. Sequence variation at cDNA positions 525, 527 and 530 have also been identified in CBA-1 and CBA-2 *Hesx1* alleles (Chapter 4). However, as the BALB/c (genomic clone) *Hesx1* allele was identical to the 129/Sv (cDNA clone) sequence at these positions, the genomic clone represents a novel *Hesx1* allele. The identification of distinct *Hesx1* alleles in the 129/Sv, BALB/c and CBA strains, presumably reflects the independent origin of these mouse strains (Altman and Katz, 1979).

5.3 Chromosomal Localization of the *Hesx1* Locus

The chromosomal localization study was carried out with a cDNA probe derived from a 950 bp *Sma* I/*Eco* RV fragment of the Hesx1a cDNA clone. It should be noted that this probe fragment contains 82 bp of sequence which is not derived from the *Hesx1* gene (see 4.2.1). At the time this experiment was carried out, it was not known that the 5' Hesx1a cDNA sequence was artifactual. However, the detection of a single intense signal using this probe argues strongly that the *Hesx1* locus was identified.

Physical mapping of the *Hesx1* locus was determined by *in situ* hybridization of mitotic chromosomes from splenic lymphocytes of BALB/c strain mice. Grains on the G-banded chromosomes were scored initially onto a standard mouse idiogram (Nesbit and Francke, 1973) and later onto a more recent idiogram of Chromosome 14 (Nadeau and Cox, 1992). A total of 159 grains over approximately 80 cells was scored over the whole karyotype. Of these, 91 grains (57% of the total) were over Chromosome 14, with the tallest column, of 50 grains, lying over band 14B and the next tallest column, of 16 grains, over sub-band A3 (Fig. 5.5A). There was no notable subpeak on the other chromosomes.

To confirm the mapping, grains were scored from approximately 70 high quality, mainly prophasic cells, with 50% more detail, onto the idiogram of Chromosome 14 of Nadeau and Cox (1992). Of a total of 103 grains, 73 (71%) formed the three tallest columns

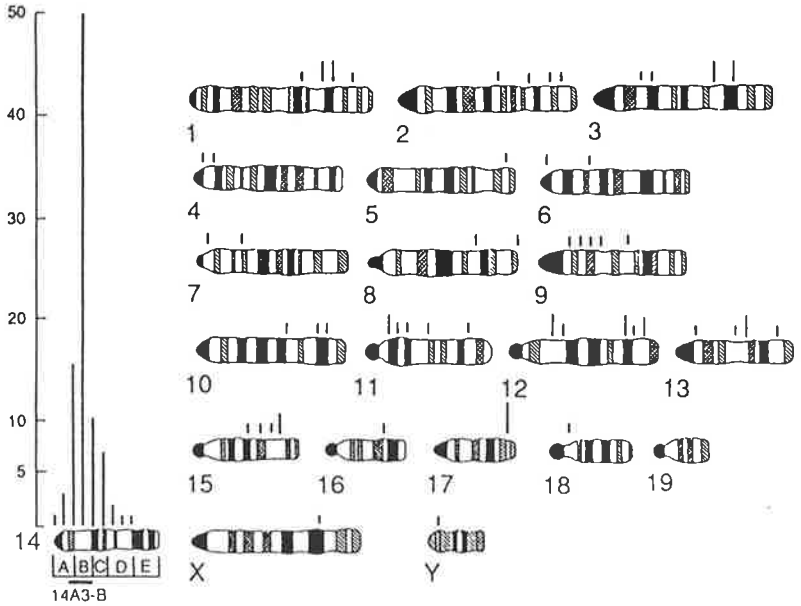
Figure 5.5 Chromosomal localization of the *Hesx1* locus.

Physical mapping of *Hesx1* was carried out using a 950 bp *Sma* I/*Eco* RV fragment derived from the Hesx1a cDNA clone.

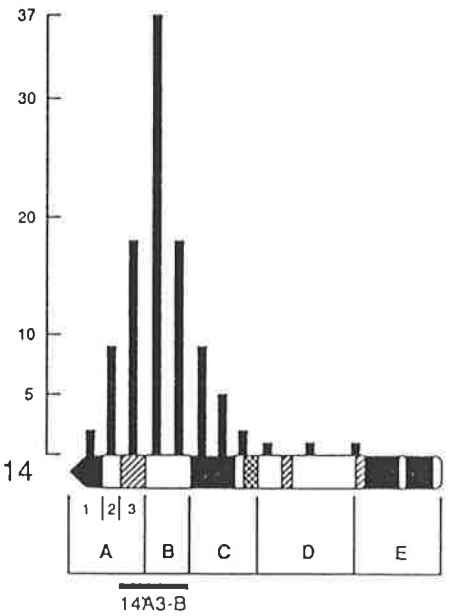
A. Grains were scored over all chromosomes from approximately 80 metaphases. 57.2% of the grains are over Chromosome 14 with the two tallest peaks over the segment 14A3-B.

B. Grains were scored to Chromosome 14 from approximately 70 high-quality, usually prophasic cells. The three tallest peaks contain 70.9% of the grains and confirm the probable location of *Hesx1* to Chromosome 14 bands A3-B.

A


50
40
30
20
10
5
14
A B C D E
14A3-B
1
2
3
4
5
6
7
8
9
10
11
12
13
15
16
17
18
19
X
Y

B


37
30
20
10
5
14
1 2 3
A
B
C
D
E
14A3-B

lying over bands 14A3-B (Fig. 5.5B), confirming these bands to be the probable location of the murine *Hesx1* gene. With both forms of scoring there was a very tall peak of grains over the proximal half of band 14B, which is regarded as the possible point location of the *Hesx1* gene.

Alignment of the physical location of *Hesx1* on chromosome 14B with the corresponding region of the chromosome 14 linkage map, predicted that *Hesx1* is located within a 7.8 cM region between the surfactant associated protein 1 gene (*Sftp-1*; Glasser *et al.*, 1990; Moore *et al.*, 1992) at 20.7 cM and the T-cell receptor alpha chain gene (*Tcra;* Kranz *et al.*, 1985) at 28.5 cM (Fig. 5.6).

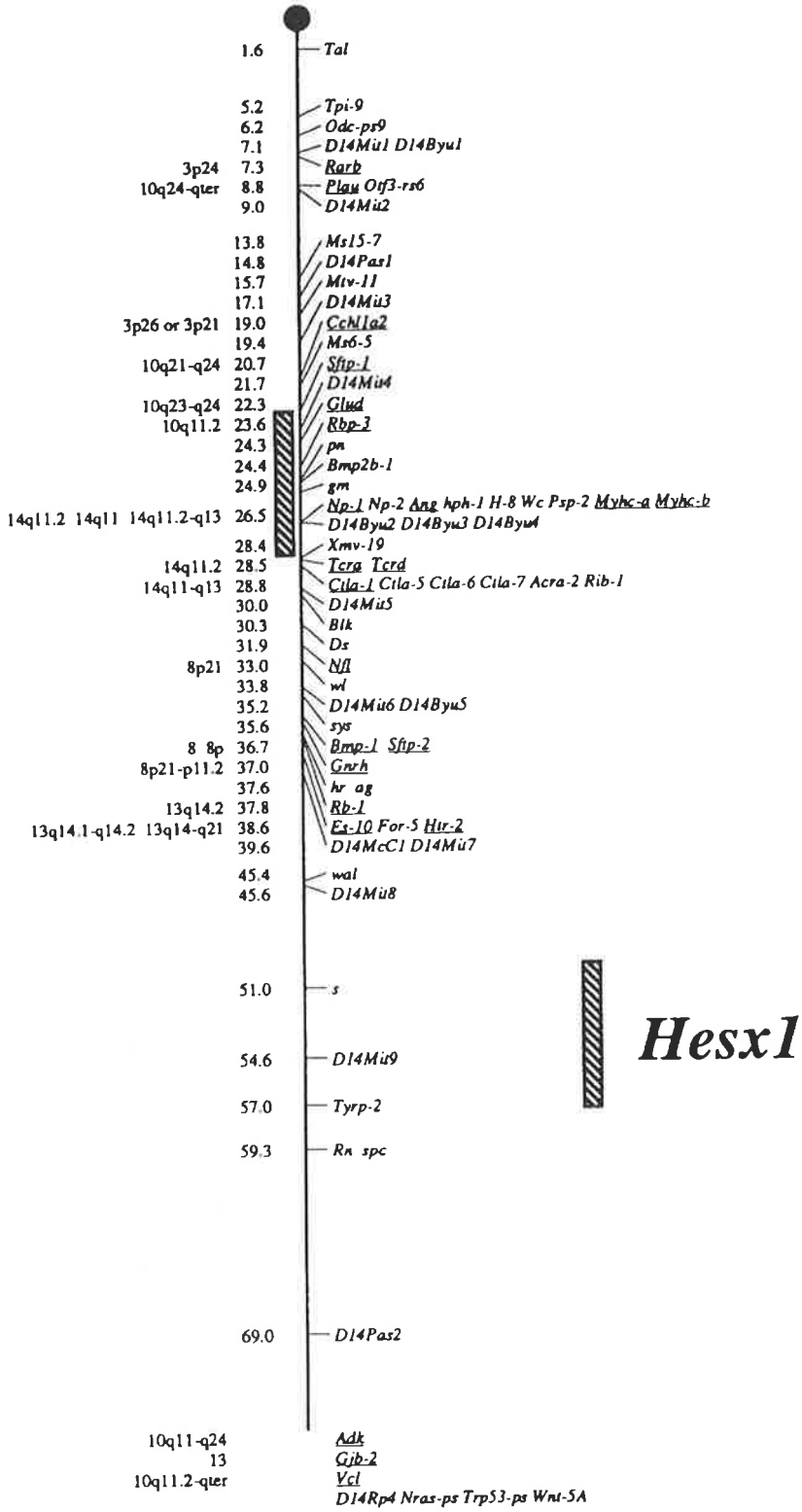
The Waved coat (*Wc*) semi-dominant developmental mutation (Diwan and Stevens, 1974) at 26.5 cM is located towards the distal end of the chromosomal region likely to contain the *Hesx1* locus (Fig. 5.6). The *Wc* mutation is named after the heterozygous phenotype which is detected at 4 to 5 days of age with the appearance of wavy irregular whiskers and short fuzzy hair on the head and upper back. *Wc* homozygous embryos die *in utero* (Green, 1972). During formation of the embryonic (primitive) and extra-embryonic ectoderm after blastocyst implantation, the embryonic ectoderm of *Wc/Wc* embryos becomes pycnotic and degenerates. This results in the formation of a "extra-embryonic organism", completely devoid of embryo proper, and consisting of a relatively normal yolk sac, chorion and allantois. The *Wc* phenotype, therefore, results from an apparent inability of the pluripotent cells to pursue normal embryonic differentiation pathways during early development.

The expression of *Hesx1 in vitro* by undifferentiated ES cells (Chapters 3 and 4), coupled with the downregulated expression in differentiated ES cells, suggested that *Hesx1* expression is likely to be restricted to the pluripotent inner cell mass and primitive ectoderm cells during early development. Given the established role of homeobox genes in the control of cellular differentiation during development (see 1.7), the apparent correlation between the *Wc* phenotype and *Hesx1* embryonic expression raised the possibility that mutation at the *Hesx1* locus may be responsible for the *Wc* phenotype. Therefore, a molecular analysis of

Figure 5.6 Location of *Hesx1* on the Chromosome 14 linkage map.

The *Hesx1* locus is located within a 7.8 cM region between the surfactant associated protein 1 gene (*Sftp-1*) at 20.7 cM and the T-cell receptor alpha chain gene (*Tcra*) at 28.5 cM. Syntenic regions in the human genome are shown on the left.

The linkage map was reproduced from Nadeau and Cox (1992).


1.6 Tal
5.2 Tpi-9
6.2 Odc-ps9
7.1 D14Mit1 D14Byu1
3p24 7.3 Rarb
10q24-qter 8.8 Plau Otf3-rs6
9.0 D14Mit2
13.8 Ms15-7
14.8 D14Pas1
15.7 Mtv-11
17.1 D14Mit3
3p26 or 3p21 19.0 Cchl1a2
19.4 Ms6-5
10q21-q24 20.7 Sftp-1
21.7 D14Mit4
10q23-q24 22.3 Glud
10q11.2 23.6 Rbp-3
24.3 pn
24.4 Bmp2b-1
24.9 gm
14q11.2 14q11 14q11.2-q13 26.5 Np-1 Np-2 Ang hph-1 H-8 Wc Psp-2 Myhc-a Myhc-b
D14Byu2 D14Byu3 D14Byu4
28.4 Xmv-19
14q11.2 28.5 Tcra Tcrd
14q11-q13 28.8 Cila-1 Cila-5 Cila-6 Cila-7 Acra-2 Rib-1
30.0 D14Mit5
30.3 Blk
31.9 Ds
8p21 33.0 Nfl
33.8 wl
35.2 D14Mit6 D14Byu5
35.6 sys
8 8p 36.7 Bmp-1 Sftp-2
8p21-p11.2 37.0 Gnrh
37.6 hr ag
13q14.2 37.8 Rb-1
13q14.1-q14.2 13q14-q21 38.6 Es-10 For-5 Htr-2
39.6 D14McC1 D14Mit7
45.4 wal
45.6 D14Mit8
51.0 s
54.6 D14Mit9
57.0 Tyrp-2
59.3 Rn spc
69.0 D14Pas2
10q11-q24 Adk
13 Gjb-2
10q11.2-qter Vcl
D14Rp4 Nras-ps Trp53-ps Wnt-5A
Hesx1

the *Hesx1* locus in *Wc* mutant mice was undertaken to determine if *Hesx1* and *Wc* were allelic.

5.4 Molecular Characterization of the *Hesx1* Locus in Waved coat (*Wc*) Heterozygous Mice

If mutation at the *Hesx1* locus is responsible for the *Wc* mutation, then the homozygous *Wc* mutant phenotype may result from inappropriate expression of *Hesx1* during early development, or alternatively, deletion or alteration of the *Hesx1* gene product. Biochemical analysis of the *Wc* homozygous embryo is extremely difficult due to the degeneration of the embryonic tissues during early development. Therefore, it was not possible to undertake analysis of *Hesx1* expression in *Wc* homozygous embryos. To address the second possibility, molecular analysis of the *Hesx1* locus in was carried out using genomic DNA derived from *Wc*/+ and +/+ littermate strain B6C3Fe-a/a-Wc mice (purchased from the Jackson Laboratory, Bar Harbor, Maine). Genomic DNA derived from *Wc* homozygous embryos was not available. Comparison between the *Hesx1* locus in the *Wc*/+ and +/+ littermate genomic DNA was undertaken using sequence and southern blot analysis.

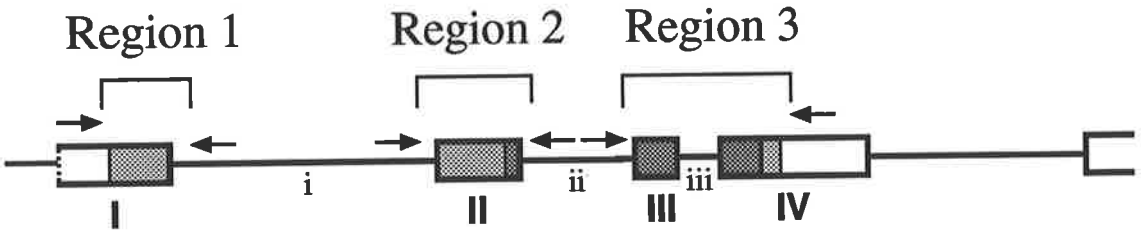
5.4.1 Sequence Analysis of the *Hesx1* Open Reading Frame (ORF)

Three regions which spanned the entire *Hesx1* ORF (which was identical in the 1.0 kb and 1.2 kb *Hesx1* transcripts, Chapter 4) were amplified from *Wc*/+ and +/+ littermate genomic DNA using PCR (Fig. 5.7). The exact positions of the primer pairs are shown in Figure 5.4, and are shown schematically in Fig. 5.7A. Region 1 primers (*Wc* 1/5 and *Wc* 1/3) amplified *Hesx1* ORF sequence within exon I, region 2 primers (*Wc* 2/5 and *Wc* 2/3) amplified *Hesx1* exon II, and the remaining ORF sequence encoded by exons III and IV was amplified by the region 3 primers (*Wc* 3/5 and *Wc* 3/3). Primers complementary to intronic sequences were designed to amplify splice doner and acceptor sites at intron/exon boundaries. As mutation within a splice doner site was identified in the Sey^{Neu} allele of the *Small eye* (*Sey*) developmental mutation (Hill *et al.*, 1991), identification of the *Hesx1*

Figure 5.7 Summary of the *Hesx1* open reading frame analysis in *Wc*/+ and +/+ genomic DNA.

A. Schematic representation of the three genomic regions amplified by the *Wc* PCR primers which encompass the entire *Hesx1* ORF. The boxed regions represent exons, and the ORF and homeobox are indicated by the light and dark shading, respectively. **B.** The *Wc*/+ and +/+ columns show the nucleotide substitutions detected in the region 1, 2 and 3 PCR products isolated from *Wc*/+ and +/+ genomic DNA. The proportion of clones which contained the altered nucleotide sequence are shown in parentheses. - indicates that all of the clones were identical to the *Hesx1* sequence shown in Fig. 4.2. The ORF column shows the position of the amino acid substitution induced by alteration of the nucleotide sequence. sil. indicates a silent nucleotide substitution.

A



B

	Wc/+	+/+	ORF
Region 1	$151^{A \rightarrow G}$ (5/8)	$151^{A \rightarrow G}$ (3/5)	$51^{N \rightarrow D}$
Region 2	$267^{T \rightarrow C}$ (1/5)	-	sil.
Region 3	$399^{T \rightarrow G}$ (1/7)	-	sil.

splice sites would indicate whether or not a similar mutation had occurred at the *Hesx1* locus in the *Wc* mutant strain. PCR products from each of the three primer pairs were digested with *Eco* RI, cloned into *Eco* RI linearized pT7T319U and sequenced in both orientations. At least five clones from each ORF region amplified from the *Wc*/+ genomic DNA were sequenced to ensure (with 95% confidence) that clones from the *Wc* allele had been isolated. These results are summarized in Figure 5.7B.

Eight region 1 clones derived from the *Wc*/+ genomic DNA were sequenced. Five of these clones were identical to the cDNA sequence derived from 129/Sv strain ES cells (Fig. 4.2). The three remaining clones contained an A to G substitution at position 151 which corresponds to a substitution of aspartate for asparagine at amino acid 51 of the *Hesx1* ORF. To determine whether the altered sequence was specific for the *Wc* allele, five region 1 clones amplified from the +/+ littermate genomic DNA were sequenced. Three of these clones also contained the single base substitution and the two remaining clones were identical to the 129/Sv *Hesx1* sequence. This demonstrated that the single base substitution was also present in one of the + alleles, and was therefore not associated with the *Wc* mutation.

Given that the A to G substitution was detected in multiple clones derived from the *Wc*/+ and +/+ genomic DNA, it seems likely that this substitution represents a sequence polymorphism. Furthermore, as sequence variation at position 151 has not been identified previously, this substitution represents a novel *Hesx1* allele. As the sequence polymorphism was detected in only one of the + alleles, it is unclear whether the 3 *Wc*/+ genomic clones which also carried the A to G substitution, were derived from the *Wc* or the + chromosomes.

Five region 2 clones derived from *Wc*/+ genomic DNA, and a single region 2 clone amplified from +/+ genomic DNA, were sequenced. Four of the *Wc*/+ clones and the single +/+ clone were identical to the 129/Sv *Hesx1* sequence corresponding to this region. A T to C substitution at position 267 was identified in the remaining *Wc*/+ clone. However, as this single base substitution is silent, it is unlikely to be responsible for the *Wc* phenotype.

A total of ten region 3 clones were sequenced, seven and three of which were derived from the *Wc*/+ and +/+ genomic DNA respectively. All of these clones were identical to the 129/Sv *Hesx1* sequence except for one *Wc*/+ clone which had a T to G substitution at position 399. Again, as this substitution corresponds to a silent substitution in the *Hesx1* ORF, it is unlikely to be associated with the *Wc* mutation.

Of the 6 region 2 and 10 region 3 clones which were sequenced, only a single clone from each region differed from the 129/Sv *Hesx1* sequence. The low frequency of these clones suggested that they could have been generated by PCR error, as opposed to representing additional sites of *Hesx1* sequence polymorphism. However, analysis of additional region 2 and region 3 clones would be required to confirm this.

Comparison of the splice donor and splice acceptor sequences of introns i, ii and iii derived from the *Wc*/+, +/+ and *Hesx1* genomic clones (5.3) indicated that these sequences were identical. The poor consensus alternative splice site within exon IV (Fig. 5.2A) was also unaltered. Therefore, the *Wc* mutation was not due to mutation within the *Hesx1* splice sites, resulting in an aberrant splicing event.

5.4.2 Southern Blot Analysis of the *Hesx1* Locus

Southern blot analysis of genomic DNA from *Wc* heterozygous mice was carried out to determine whether chromosomal rearrangements at the *Hesx1* locus were associated with the *Wc* mutation. 5 µg of *Wc*/+ and +/+ littermate genomic DNA was digested with *Eco* RI and *Hind* III in single and double digests, electrophoresed on a 0.8% agarose gel and blotted. 0.5 µg Hesx1g (5.2.2) was also digested to allow direct comparison between band sizes in the B6C3Fe-a/a-Wc strain genomic DNA and in the *Hesx1* genomic clone.

A 9 kb *Pst* I fragment from Hesx1g which spanned most of the *Hesx1* transcription unit and flanking sequences (Fig 5.2A) was used to probe the *Wc* genomic southern. After hybridization, the southern blot was washed in 0.2 X SSC/0.1% SDS at 25°C, 50°C and 68°C for 20 minutes per washing solution. The genomic probe generated an intense smear down each of the genomic DNA tracks (data not shown). This indicated that the 9 kb *Hesx1*

genomic fragment contained at least one genomic repeat sequence which resulted in hybridization to multiple genomic fragments.

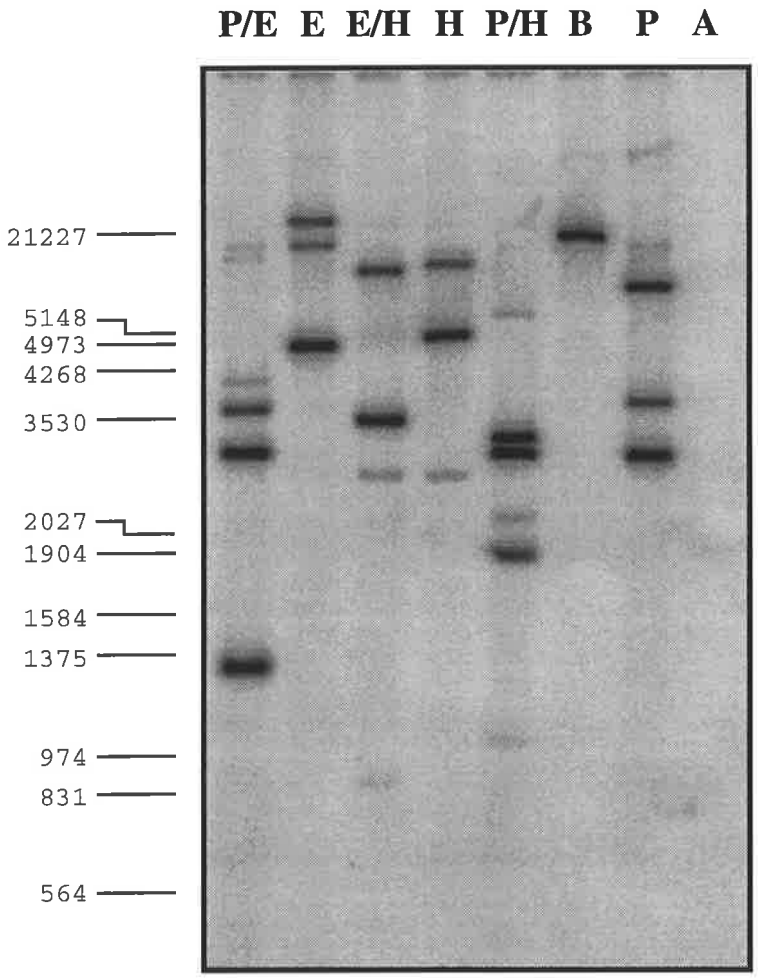
To identify regions within the *Hesx1* genomic clone which did not contain repetitive elements, and could therefore be used as *Hesx1* specific probes, a reverse southern was carried out. 1 μg of pHesx1g plasmid was digested with *Eco* RI, *Hind* III, *Bam* HI and *Pst* I in single and double digests, electrophoresed on a 1% agarose gel and blotted. As a negative (single copy gene) control, 1 μg of a pT7T319U plasmid containing a 402 bp *Alu* I fragment spanning positions 31 to 432 of the Hesx1a cDNA, was also loaded. A murine genomic probe prepared from 500 ng strain 129/Sv genomic DNA was hybridized to the southern followed by washing in 0.2 X SSC / 0.1% SDS at 25°C, 50°C and 68°C for 20 minutes per washing solution (Figure 5.8A). *Hesx1* genomic fragments which contained repetitive element sequence were identified by their hybridization to the genomic DNA probe. These regions are shown schematically in Figure. 5.8B. No signal was detected in the negative control lane, confirming that only repetitive sequences were identified by this analysis. A range of signal intensities was detected in the *Hesx1* genomic fragments which hybridized to the genomic probe. This indicated that multiple repetitive elements, which differed in their genomic frequency, were present at the *Hesx1* locus.

Genomic repeat sequences were not detected within a 3.6 kb region containing exons 1 and 2, and 2.4 kb of upstream sequence. A 1.9 kb *Hind* III fragment derived from the 5' end of this region was selected to probe the *Wc* genomic southern (Fig. 5.9). This region lies immediately upstream of the known *Hesx1* transcription unit and is therefore likely to harbour the *Hesx1* promoter. It was also possible, based on other studies of *Hox-d4* (*Hox-4.2*) and *Hox-b9* (*Hox-2.5*) homeobox genes, that this genomic region may contain sequences which regulate gene expression (Popperl and Featherstone, 1993; Kondo *et al.*, 1992). Single bands of 1.9 kb and 1.6 kb were detected in the *Hind* III and *Eco* RI/*Hind* III digests of the *Wc*/+ and +/+ genomic DNA, and in the pHesx1g plasmid (Fig. 5.9). This suggested that this region of genomic DNA is identical in the *Wc*/+ and +/+ genomic DNA (and in the *Hesx1* genomic clone). This analysis, therefore, did not support an allelic relationship between *Hesx1* and *Wc*. However, it should be noted that minor alterations at

Figure 5.8 Detection of repetitive elements at the *Hesx1* locus by reverse southern blot analysis.

A. 1 μg Hesx1g was digested, electrophoresed on a 1% agarose gel and blotted. The southern blot was hybridized to oligolabelled murine genomic DNA and washed in 0.2 X SSC / 0.1% SDS at 25°C, 50°C and 68°C for 20 minutes per washing solution. The filter was exposed overnight in a phosphorimager cassette. The bands indicate *Hesx1* genomic sequences which contain repetitive elements. No signal was detected in the single copy control lane (A) which contained 1 μg of a pT7T319U plasmid containing a 402 bp *Alu* I fragment spanning positions 31 to 432 of the Hesx1a cDNA. The position of the DNA markers is shown on the left. P=*Pst* I; E=*Eco* RI; H=*Hind* III; B=*Bam* H1. **B.** *Hesx1* genomic regions which contain repeat sequences are indicated by the striped boxes. Exons are shown as black boxes.

A


P/E E E/H H P/H B P A
21227
5148
4973
4268
3530
2027
1904
1584
1375
974
831
564

B

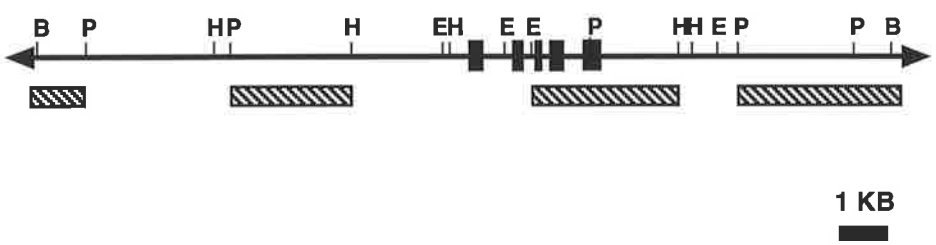
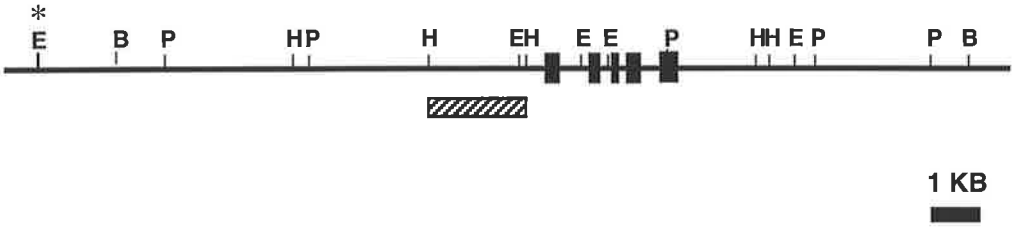

B P HP H EH E E P HH E P P B
1 KB

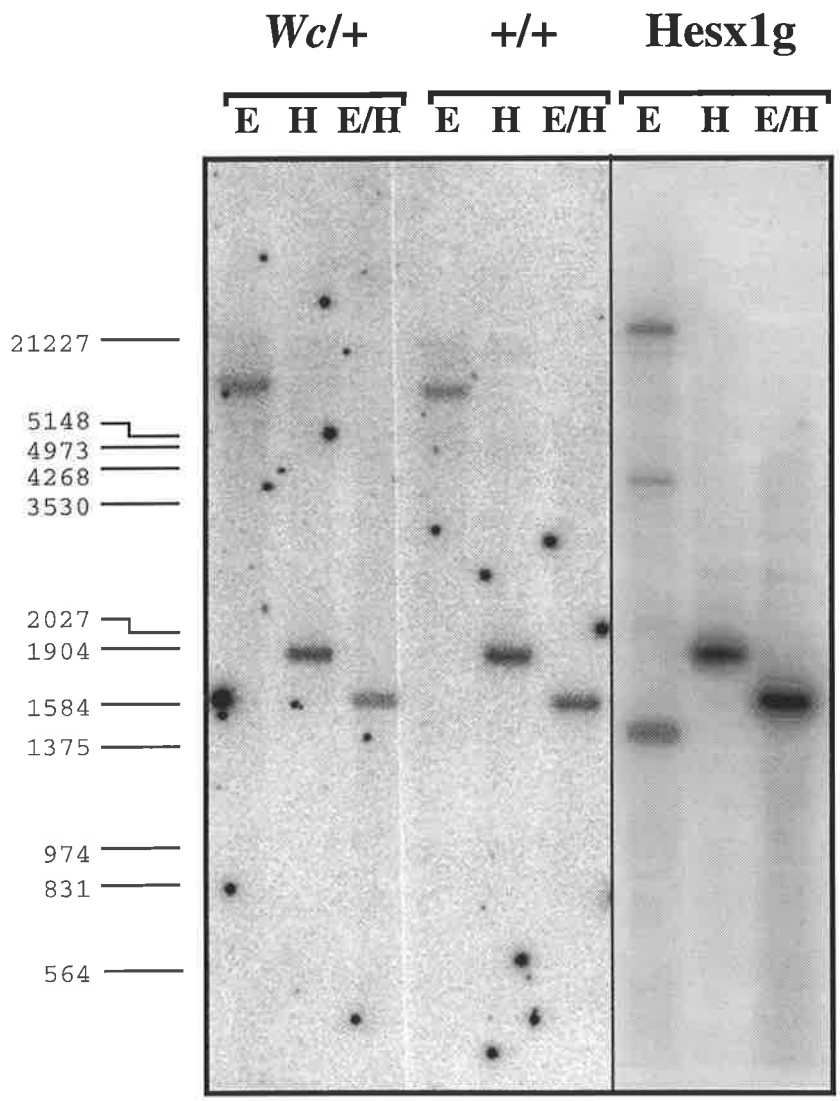
Figure 5.9 Southern blot analysis of the putative *Hesx1* promoter region in *Wc*/+ and +/+ genomic DNA.

A. Restriction map of the *Hesx1* locus. The 1.7 kb *Hind* III probe fragment is indicated by the striped box. *Hesx1* exons are shown as black boxes. The additional *Eco* RI site detected in the *Wc*/+ and +/+ genomic DNA is marked by an asterisk (see 5.4.2). P=*Pst* I; E=*Eco* RI; H=*Hind* III; B=*Bam* H1. **B.** 5 μg of *Wc*/+ and +/+ genomic DNA, and 0.5 μg of Hesx1g, was digested with *Eco* RI and/or *Hind* III, electrophoresed on a 0.8% agarose gel and blotted. The 1.7 kb *Hesx1* genomic probe was hybridized to the southern blot overnight, followed by washing in 0.2 X SSC / 0.1% SDS at 25°C, 50°C and 68°C for 20 minutes per washing solution. DNA marker sizes are shown on the left. Note that as the Hesx1g bands were of much greater intensity than the genomic DNA bands, different relative exposures of the *Hesx1* and genomic DNA lanes are shown to allow direct comparison between these tracks.

A


*
E
B
P
HP
H
EH
E E
P
HH E P
P B
1 KB

B


Wc/+
+/+
Hesx1g
E H E/H E H E/H E H E/H
21227
5148
4973
4268
3530
2027
1904
1584
1375
974
831
564

the *Hesx1* locus such as point mutations and small deletions, which may alter the regulation of *Hesx1* expression, would not have been detected in this analysis.

A single band (approximately 10 kb) was detected in the *Eco* RI digested *Wc*/+ and +/+ genomic DNA. This indicated the existence of an additional upstream *Eco* RI site (marked by the asterisk in Fig. 5.9A) which was not detected in Hesx1g as the genomic clone did not extend into this 5' region.

The *Wc* southern was also probed with a cDNA probe prepared from a 941 bp *Afl* III fragment (spanning positions -44 to 900 of Hesx1b, Fig. 4.2) which contained the *Hesx1* ORF (Fig. 5.10A,B). No difference in banding pattern between the *Wc*/+ and +/+ genomic DNA was detected, which indicated that significant rearrangement and/or deletion at the *Hesx1* locus had not occurred in the *Wc* mutant strain. Again, it should be noted that minor differences in the +/+ and *Wc*/+ genomic DNA as a result of small deletions or point mutations, may not have been detected by this analysis.

Several differences between the B6C3Fe-a/a-Wc genomic DNA and the Hesx1g (BALB/c) genomic clone were identified:

a) The 1.4 kb *Eco* RI band was absent from the *Wc*/+ and +/+ genomic digests.

b) The 5.1 kb *Hind* III fragment in the genomic clone had increased to approximately 5.9 kb in the genomic DNA tracks.

c) Two bands of 3 kb and 1.9 kb were detected in the genomic DNA after *Eco* RI/*Hind* III double digestion .

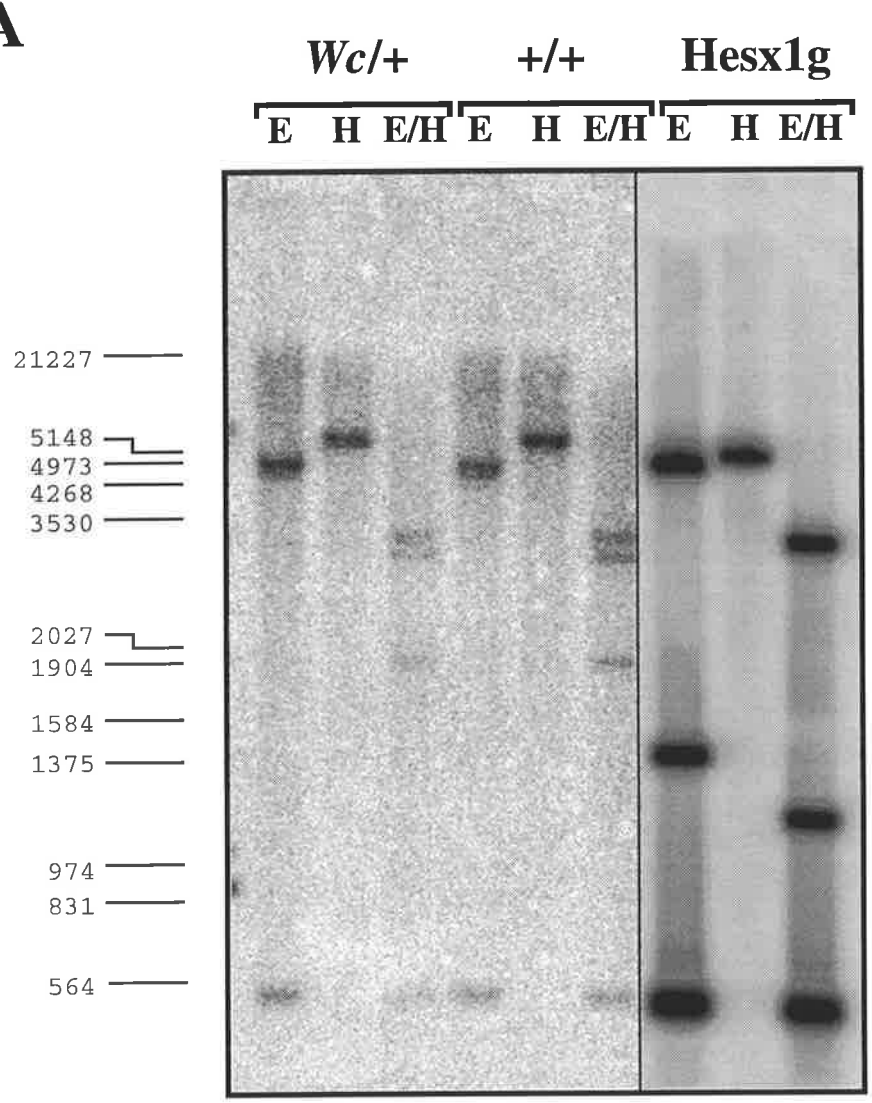
d) The 1.1 kb *Eco* RI/*Hind* III fragment in the genomic clone was not present in the genomic DNA samples.

One possible explanation for this polymorphism is shown in Fig. 5.10C. The detection of the 1.6 kb *Eco* RI/*Hind* III band (H-Ea) and the 1.9 kb *Hind* III band (H-Ha) with the 1.9 kb genomic probe (Fig. 5.9), and the 500 bp *Eco* RI band (Eb-Ec) with the cDNA probe (Fig. 5.10A), indicated that the H, Ea, Ha, Eb and Ec sites were present in the B6C3Fe-a/a-Wc genomic DNA. Therefore, the absence of a 1.1 kb band corresponding to the region between Ha and Eb, and the detection of a 1.9 kb band in the *Eco* RI/*Hind* III

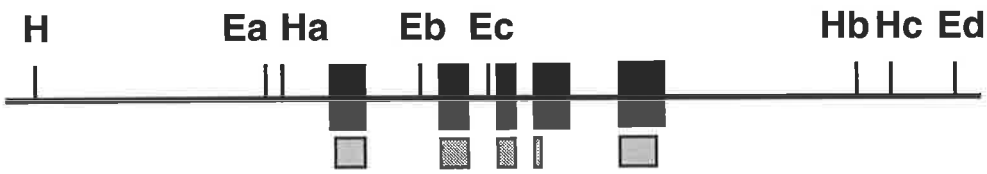
Figure 5.10 Southern blot analysis of the *Hesx1* open reading frame in *Wc*/+ and +/+ genomic DNA.

A. The genomic southern blot (see Fig. 5.9) was probed overnight with an oligolabelled probe derived from a 941 bp *Afl* III cDNA fragment (spanning positions -44 to 900 of Hesx1b) which contained the entire *Hesx1* open reading frame. The filter was washed in 0.2 X SSC / 0.1% SDS at 25°C, 50°C and 68°C for 20 minutes per washing solution. DNA marker sizes are shown on the left. Note that as the Hesx1g bands were of much greater intensity than the genomic DNA bands, different relative exposures of the *Hesx1* and genomic DNA lanes are shown to allow direct comparison between these tracks. **B.** Restriction map of *Hesx1* showing the genomic regions which bind the ORF probe (shaded boxes). The exons are indicated by the black boxes. **C.** Restriction map of *Hesx1* showing the approximate location of the proposed insert event (which lies between Ha and the 5' end of exon I) and the position of the additional *Hind* III site (H'). The Hb site which is proposed to be absent from one of the genomic *Hesx1* alleles is marked by an asterisk.

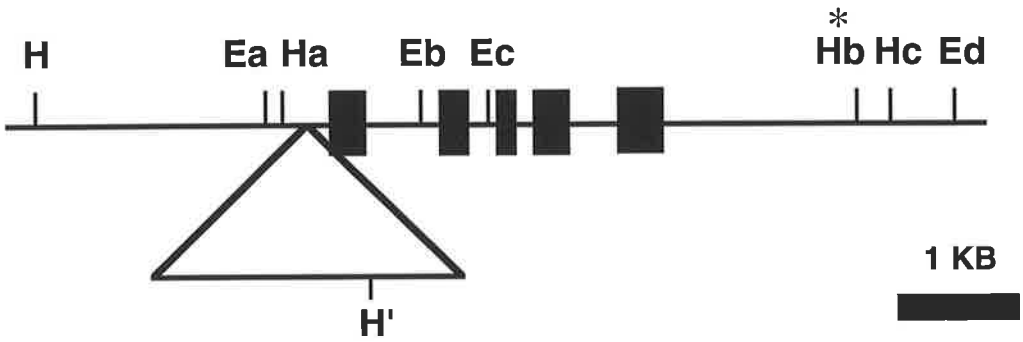
A


Wc/+
+/+
Hesx1g
E H E/H E H E/H E H E/H
21227
5148
4973
4268
3530
2027
1904
1584
1375
974
831
564

B


H
Ea Ha
Eb Ec
Hb Hc Ed

C


H
Ea Ha
Eb Ec
*
Hb Hc Ed
H'
1 KB

digested B6C3Fe-a/a-Wc genomic DNA, suggested that additional DNA was present between these restriction sites. This may be explained by an insertion event between the Ha and Eb sites which introduced an additional 2.6 kb (approximately) of DNA (Fig. 5.10C). This inserted DNA contained an additional *Hind* III site (H') 1.9 kb upstream of the Eb site, which gives rise to the 1.9 kb band detected in the *Eco* RI/*Hind* III digested genomic DNA. The 2.6 kb insertion predicts that the 1.4 kb *Eco* RI band (between Ea and Eb, Fig. 5.10B), which was not detected in the *Wc* genomic DNA, would become 4 kb in length. This 4 kb *Eco* RI band in the *Eco* RI tracks may be masked by the 4.3 kb band (corresponding to the region between Ec and Ed). The detection of the additional 3.3 kb band in the *Eco* RI/*Hind* III track may be due to the absence of the Hb (*Hind* III) site in one of the *Hesx1* alleles. This was supported by the similar intensities of the 3 kb and the 3.3 kb bands resulting from *Eco* RI/*Hind* III digestion of the B6C3Fe-a/a-Wc genomic DNA. The absence of the Hb site was not detected in the *Hind* III digest because the 5.9 kb (between Ha and Hc) band masked the 5.6 kb signal.

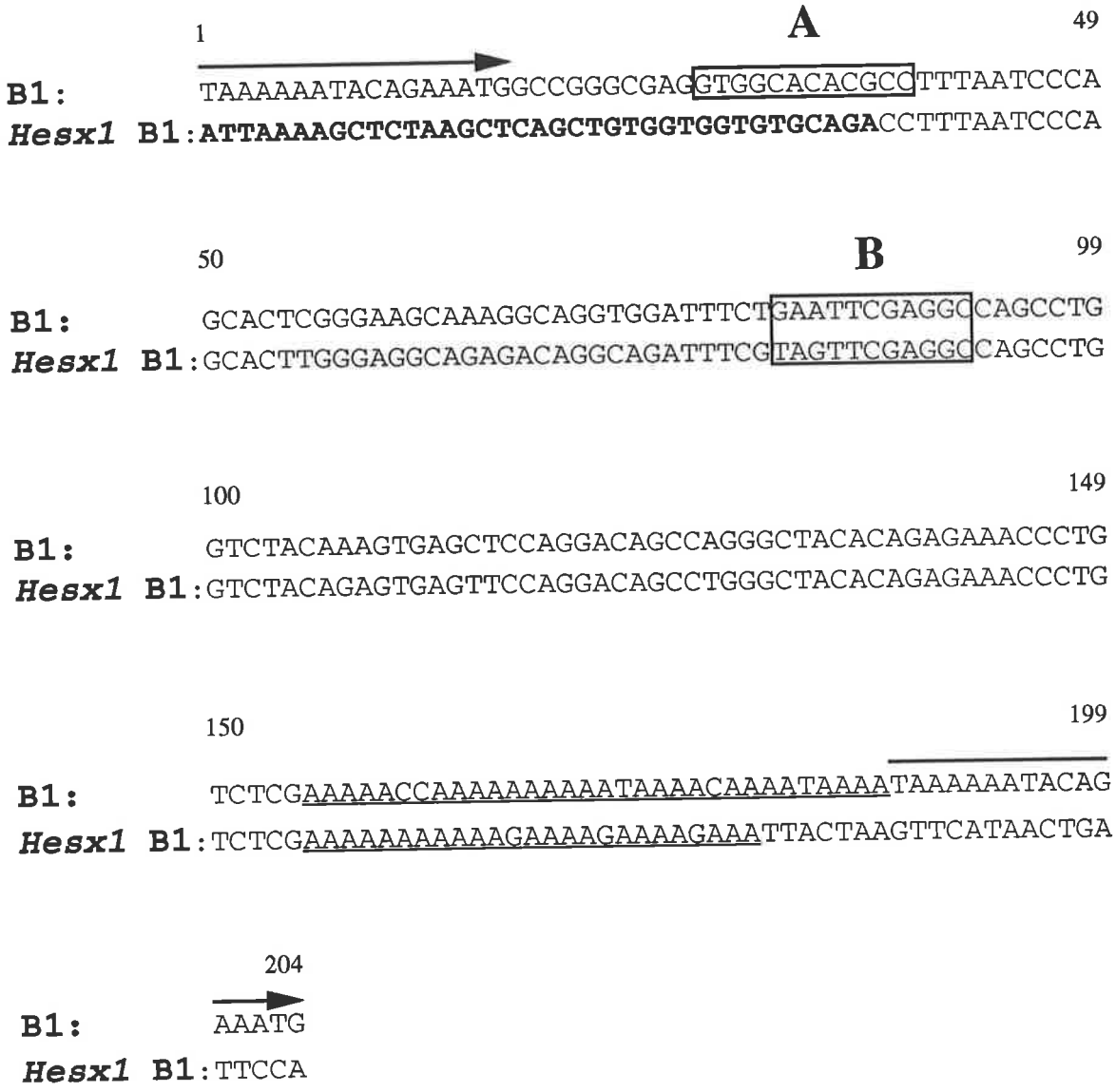
5.5 Identification of a Genomic Repeat Sequence at the *Hesx1* Locus

The reverse southern blot analysis of the *Hesx1* genomic region indicated that the 1.3 kb *Eco* RI/*Pst* I fragment spanning exons III and IV and part of exon V contained a genomic repeat sequence (Fig. 5.8A,B). As genomic repetitive elements have been shown to regulate expression of neighbouring genes, the identity of this putative genomic repeat sequence was investigated. Intronic sequence within the 1.3 kb *Eco* RI/*Pst* I fragment included the 19 3' nucleotides of intron ii, and the entire intron iii and iv sequences. Comparison of the intron iii sequence to the EMBL and GenBank databases sequence did not identify any genomic repeat sequences, and therefore it seemed probable that intron iv harboured the repetitive element. Intron iv was sequenced using oligonucleotide primers (rep5 and rep3) complimentary to the flanking exon IV and exon V sequences (Fig. 5.4), and compared with the EMBL and GenBank databases.

A 167 bp B1 repeat sequence (corresponding to bases 2359 to 2498 of the *Hesx1* genomic sequence, Fig. 5.4) was identified within intron iv (Fig. 5.11). The B1 repeat

Figure 5.11 Nucleotide sequence of the *Hesx1* B1 repeat.

Alignment of a B1 repeat sequence (Hastie, 1989) with the *Hesx1* B1 repeat. The A and B boxes of the RNA polymerase III promoter are boxed. The terminal repeat sequences are indicated by arrows and the $(A)_n$ tail is underlined. The 5' *Hesx1* sequence which did not correspond to the B1 repeat sequence is shown in bold.


1
A
49
B1: TAAAAAATACAGAAATGGCCGGGCGAGGTGGCACACGCCTTTAATCCCA
Hesx1 B1: ATTAAAAGCTCTAAGCTCAGCTGTGGTGGTGTGCAGACCTTTAATCCCA
50
B
99
B1: GCACTCGGGAAGCAAAGGCAGGTGGATTTCTGAATTCGAGGCCAGCCTG
Hesx1 B1: GCACTTGGGAGGCAGAGACAGGCAGATTTCGTAGTTCGAGGCCAGCCTG
100
149
B1: GTCTACAAAGTGAGCTCCAGGACAGCCAGGGCTACACAGAGAAACCCTG
Hesx1 B1: GTCTACAGAGTGAGTTCCAGGACAGCCTGGGCTACACAGAGAAACCCTG
150
199
B1: TCTCGAAAAACCAAAAAAAAAATAAAACAAAATAAAATAAAAAATACAG
Hesx1 B1: TCTCGAAAAAAAAAAGAAAAGAAAAGAAATTACTAAGTTCATAACTGA
204
B1: AAATG
Hesx1 B1: TTCCA

belongs to the *Alu* family of SINE (short interspersed element) repetitive elements, and is believed to be derived from the highly conserved 7SL RNA (Hastie, 1989). The 7SL RNA has a key role in the processing of proteins destined to be transported from the cell (Walter and Blobel, 1982). B1 repeats are the most highly repeated dispersed element in the murine genome, and are repeated 1.3-1.8 X 10^5 times, constituting 0.7-1.0% of the mouse genome (Hastie, 1989; Bennett *et al.*, 1984). The intensity of the band corresponding to the 1.2 kb *Eco* RI/*Pst* I genomic fragment containing the B1 repeat (Fig. 5.8A), and other larger fragments which also contained this region, was consistent with the great abundance of this repetitive element within the murine genome. Although B1 repeats have been detected within many genes, sometimes in multiple copies (Bourbon *et al.*, 1988), these repetitive elements do not appear to play a role in gene regulation (Saksela and Baltimore, 1993).

Three structural features which are found within the B1 repeat are a $(A)_n$ tail, short flanking direct repeat sequences, and the A and B boxes of the RNA polymerase III promoter sequence which are derived from the 7SL RNA sequence (Fig. 5.11). The *Hesx1* B1 repeat contained a 25 bp imperfect $(A)_n$ tail, and the RNA polymerase III B box, but was unusual in that the RNA polymerase III A box was not present (Fig. 5.11). In addition, the short flanking direct repeat sequences could not be identified. This suggested that approximately 37 bp at the 5' end of the *Hesx1* B1 repeat was missing. It is not clear whether this 5' sequence was lost during insertion of the repetitive element, or as a post-insertional deletion event.

5.6 Discussion

A 16 kb *Hesx1* genomic clone derived from a BALB/c strain genomic library was isolated and the architecture of the *Hesx1* gene was determined by comparison of the cDNA and genomic sequences. Four introns within the *Hesx1* gene were identified. Although single introns within the homeobox occur in many homeobox genes from a variety of species, the *Hesx1* genomic structure was unusual in that two of these introns (ii and iii) were located within the homeobox. This relatively rare arrangement has not been previously described in vertebrates (Burglin, 1994). The position of the first intron was shared with the

homeobox gene *mec-3* (Way and Chalfie, 1988), while the location of the downstream intron, within helix 3 of the homeobox, was shared with a variety of homeobox genes (Burglin, 1994). As the genomic architecture of the *Xenopus* homeobox gene XANF-1 (Zariasky *et al.*, 1992) has not been described, it remains to be determined whether this structural feature is conserved within the *Hesx1*/XANF-1 class of homeobox genes.

Correlation between the position of exonic sequences, and the existence of discrete domains within protein structure, has led to the proposition that exons encode functional domains or modules, which may used to assemble additional related genes during evolution (Gilbert, 1985). As the homeobox encodes a functional DNA binding domain (the homeodomain), which is conserved throughout evolution, it might be expected that in most cases, the homeobox would be contained within a single exon. However, the regular occurrence of introns within the homeobox indicates that the homeobox is not generally encoded by a single exon, and is therefore not consistent with the domain theory of intron/exon organisation. One possible explanation for the presence of introns within the homeobox is that the intronic sequences harbour regulatory elements which control gene expression.

The molecular difference between the 1.0 kb and 1.2 kb *Hesx1* transcripts was shown to arise from differential usage of a poor consensus splice site located immediately downstream of the open reading frame termination codon. The biological role of the alternative *Hesx1* transcripts, which encode identical proteins, remains obscure. However, given that expression of the 1.2 kb *Hesx1* transcript has been detected in ES cells but not in RNA derived from 14.5 days p.c. and 16 days p.c. embryos (Chapter 4), it is possible that differential splicing of the *Hesx1* primary transcript is developmentally regulated. Furthermore, the apparent restriction of the 1.2 kb *Hesx1* transcript to undifferentiated ES cells may reflect a unique developmental function for this transcript in the pluripotent cells of the early murine embryo. This may be investigated further by *in situ* hybridization analysis of the early embryo using transcript specific probes.

In situ hybridization of a *Hesx1* cDNA probe identified the probable location of the *Hesx1* locus as chromosome 14 bands A3-B. This location does not correspond to any of

the 4 *Hox* gene clusters in the mouse, which is consistent with the lack of homology between *Hesx1* and the *Hox* genes (Chapter 4). The absence of other known homeobox genes within this region is consistent with studies which show that homeobox genes which are not part of the *Hox* clusters are scattered throughout the genome.

Chromosomal localization of developmental control genes has proved useful in identifying the molecular basis for known developmental mutations (Gruss and Walther, 1992). The identification of the early developmental mutation Waved coat (*Wc*) within the predicted *Hesx1* chromosomal location, prompted a molecular investigation of the *Hesx1* ORF and genomic region in *Wc*/+ and +/+ genomic DNA. Comparison of the *Hesx1* open reading frame sequences generated from *Wc*/+ and +/+ genomic DNA with the *Hesx1* cDNA sequence (Fig. 5.7B), identified three single nucleotide differences. An A to G substitution at position 151 of the ORF, which resulted in an asparagine to aspartate substitution at position 51, was identified in 3 of 8 and 3 of 5 *Wc*/+ and +/+ clones respectively. This sequence variation was detected in the *Wc*/+ and +/+ clones, and was therefore not specific to the *Wc* allele. As the genomic DNA clones and the *Hesx1* cDNA clones were derived from different strains, it is probable that this sequence variation was due to polymorphism. Silent base changes at positions 267 and 399 of the open reading frame were also detected in clones derived from the *Wc*/+ genomic DNA. As these substitutions did not alter the identity of the respective codons, it seems very unlikely that they are associated with the *Wc* mutation. In conclusion, sequence analysis of the *Hesx1* ORF did not detect sequence variation which was specific for the *Wc* allele, and therefore excluded the possibility that mutation within the *Hesx1* open reading frame had resulted in the *Wc* mutation.

The possibility that more extensive mutation at the *Hesx1* locus had occurred in the *Wc* mutant was investigated by southern blot analysis of *Wc*/+ and +/+ genomic DNA. Hybridization of a *Hesx1* upstream genomic fragment, and a cDNA fragment containing the open reading frame, failed to detect any difference in banding patterns between the *Wc*/+ and +/+ genomic DNA. This indicated that significant mutation, such a deletion or insertion event, had not occurred at the *Hesx1* locus on chromosomes carrying the *Wc* mutation.

However, the possibility that subtle mutation such as single base changes (outside the open reading frame) or small deletions or insertions had occurred could not be excluded by this analysis.

Interestingly, a difference in the structure of the *Hesx1* locus in the B6C3Fe-a/a-Wc genomic DNA and the Hesx1g (BALB/c) genomic clone was detected. This polymorphism was proposed to be due to the insertion of approximately 2.6 kb of DNA, between the Ha and Eb restriction sites (Fig. 5.10C). The detection of this additional DNA in the wild-type (+/+) genomic DNA suggests that the insertion event had not resulted in the mutation responsible for the *Wc* phenotype. Although the origin of this additional DNA remains obscure, one possibility is that it may be due to insertion of a retrotransposon, perhaps belonging to the LINE (long interspersed element) family of repetitive elements (Singer, 1982).

In summary, sequence and southern blot analysis of *Hesx1* in *Wc*/+ and +/+ strain genomic DNA did not support an allelic relationship between *Hesx1* and *Wc*. However, it remains possible that mutation at the *Hesx1* locus, such as alteration of promoter elements, which may result in the deregulation of *Hesx1* expression, may have occurred in the *Wc* strain. This possibility may be investigated by *in situ* or RNAase protection analysis of *Hesx1* expression in the inner cell mass cells of *Wc* homozygous embryos. Further mapping studies may also be useful to determine the relationship between *Hesx1* and the *Wc* mutation. The 7.8 cM region on chromosome 14 bands A3-B which define the *Hesx1* location, is syntenic proximally and distally on human 10q and human 14q, respectively (Nadeau and Cox, 1992; Fig. 5.6). Given that the region around the *Wc* locus is syntenic on human 14q11, physical location of a human *Hesx1* homolog to chromosome 14 would support the assignment of the *Wc* mutation to this locus.

A 139 bp B1 repeat sequence was detected within intron iv of the *Hesx1* gene. Although B1 repeat sequences have been identified within many genes including alpha-fetoprotein (Young *et al.*, 1982), nucleonin (Bourbon *et al.*, 1988) and glycerophosphate dehydrogenase (Phillips *et al.*, 1986), a functional role for these repetitive elements in gene regulation has not been described (Schmid and Jelinek, 1982). Interestingly, RNA

polymerase III mediated transcription of B1 repeats has been shown to occur *in vivo* (Adeniyi-Jones and Zasloff, 1985). Furthermore, the distribution of the B1 transcripts appears to be tissue specific (Adeniyi-Jones and Zasloff, 1985), although the biological significance of this is not known. However, given that the RNA polymerase III A box in the *Hesx1* B1 repeat was not present, it seems likely that the *Hesx1* repeat is transcriptionally inactive.

CHAPTER 6: ANALYSIS OF *Hesx1* OVER-EXPRESSION IN EMBRYONIC STEM CELLS

CHAPTER 6: ANALYSIS OF *Hesx1* OVER-EXPRESSION IN EMBRYONIC STEM CELLS

6.1 Introduction

Comparison between *Hesx1* and *Oct-4* expression in undifferentiated ES cells and chemically- and spontaneously-differentiated ES cells, indicated that *Hesx1* expression was downregulated during ES cell differentiation (Chapter 3). *Hesx1* expression *in vitro* therefore suggests that in the pre-gastrulation embryo, *Hesx1* would be expressed by pluripotent cells, and that expression would be downregulated during cellular differentiation. Given that homeobox genes have been shown to play a role in the specification of cellular identity (1.7), *Hesx1* expression *in vitro* suggests that this gene may play a role in the maintenance of the pluripotent cell phenotype *in vivo*. Alternatively, *Hesx1* may play a part in lineage specification during pluripotent cellular differentiation. The developmental role of *Hesx1* would presumably be mediated by specific interaction between the *Hesx1* homeodomain and regulatory elements of downstream target genes.

One approach toward the investigation of the developmental role of *Hesx1* in pluripotent cells, is to characterize the effect of *Hesx1* over-expression on ES cell differentiation. If *Hesx1* plays a role in the maintenance of pluripotency, then *Hesx1* over-expression may result in the inhibition of cellular differentiation. As ES cells are dose-responsive to the cytokine LIF for maintenance of the undifferentiated state (Smith, 1991), direct analysis of the effect of *Hesx1* over-expression on ES cell differentiation can be carried out by culture of *Hesx1* over-expressing ES cell lines in a range of LIF concentrations. In addition, the possible role of *Hesx1* in lineage specification during pluripotent cell differentiation may be investigated by analysis of the formation and differentiation of embryoid bodies (1.4.3.2) derived from *Hesx1* over-expressing ES cell lines. This approach allows direct analysis of the *Hesx1* function during ES cell differentiation within a complex "embryo-like" environment.

Expression vectors which contain constitutive promoters are unsuitable for this type of investigation because stable ES cell transfectants cannot be generated in the absence of

transgene expression. As a consequence, direct correlation between transgene expression and cellular differentiation is not possible. Therefore, the interferon inducible vectors p1OX and p3OX (Whyatt *et al.*, 1993) were used for *Hesx1* over-expression analysis (Fig. 6.1). These vectors contain interferon-stimulable response elements (ISREs) derived from the promoter region of the human 6-16 gene (Kelly *et al.*, 1985; Kelly *et al.*, 1986), and allow regulation of transgene expression in ES cell transfectants in an interferon dose-dependent manner. These vectors are particularly useful to investigate genes which are expressed at low levels, as they direct extremely low levels of background expression. The p1OX vector, which contains a longer 6-16 promoter fragment including intron 1 sequence (Fig. 6.1), has been shown to direct greater levels of transgene expression, albeit with a detectable level of basal (uninduced) expression (Whyatt *et al.*, 1993).

This chapter describes the generation of 1OX and 3OX *Hesx1* over-expression constructs, and the subsequent analysis of *Hesx1* over-expression in uninduced and induced ES cell transfectants.

6.2 Generation and Characterization of *Hesx1* Inducible Stable ES Cell Lines

6.2.1 Generation of p1Hesx1X and p3Hesx1X Stable ES Cell Lines

RNAase protection analysis indicated that the 1.2 kb *Hesx1* transcript was expressed by undifferentiated ES cells, but was not detected in tissue specific embryonic RNA samples derived from 14.5 days p.c. and 16 days p.c. embryos (Fig. 4.8). Although the 1.2 kb and 1.0 kb are likely to encode identical open reading frames, the apparent restriction of the 1.2 kb *Hesx1* transcript to ES cells is consistent with a specific role for this transcript in pluripotent cells. Therefore, interferon inducible expression constructs for *Hesx1* over-expression analysis were generated from the Hesx1b cDNA clone (which corresponded to the 1.2 kb *Hesx1* transcript).

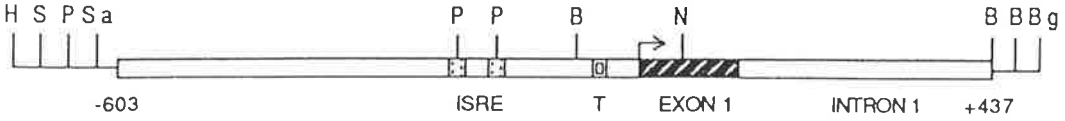
A 944 bp *Afl* III fragment (spanning positions -44 to 900 of Hesx1b (Fig. 4.1)) encoding the *Hesx1* ORF was end-filled and ligated into the *Bgl* II site in the p1OT and

Figure 6.1 Construction of the p1OX and p3OX expression vectors.

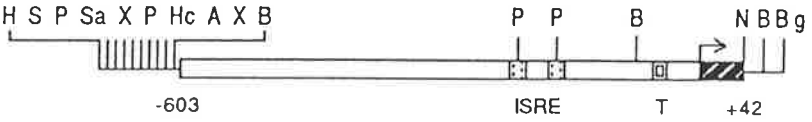
The 1040 bp (A) and 645 bp (B) fragments derived from the 6-16 gene promoter (Kelly *et al.*, 1985; Kelly *et al.*, 1986) were inserted into the pOX vector backbone to generate the p1OX and p3OX expression vectors (C). The unique *Bgl* II site into which the *Hesx1* open reading frame was inserted is indicated by the bold arrow. Abbreviations: ISRE, Interferon-stimulable response element; T, TATA box; DHFR, murine dihydrofolate reductase gene; VAI, adenovirus VAI gene; ORI, *E. coli* origin of replication; horizontal arrow (A, B), transcription start site. Restriction sites: A, *Acc* I; B, *Bam* HI; Bg, *Bgl* II; E, *Eco* RI; H, *Hind* III; Hc, *Hinc* II; N, *Nci* I; P, *Pst* I; S, *Sph* I; Sa, *Sal* I; X, *Xba* I.

Adapted from Whyatt *et al.* (1993).

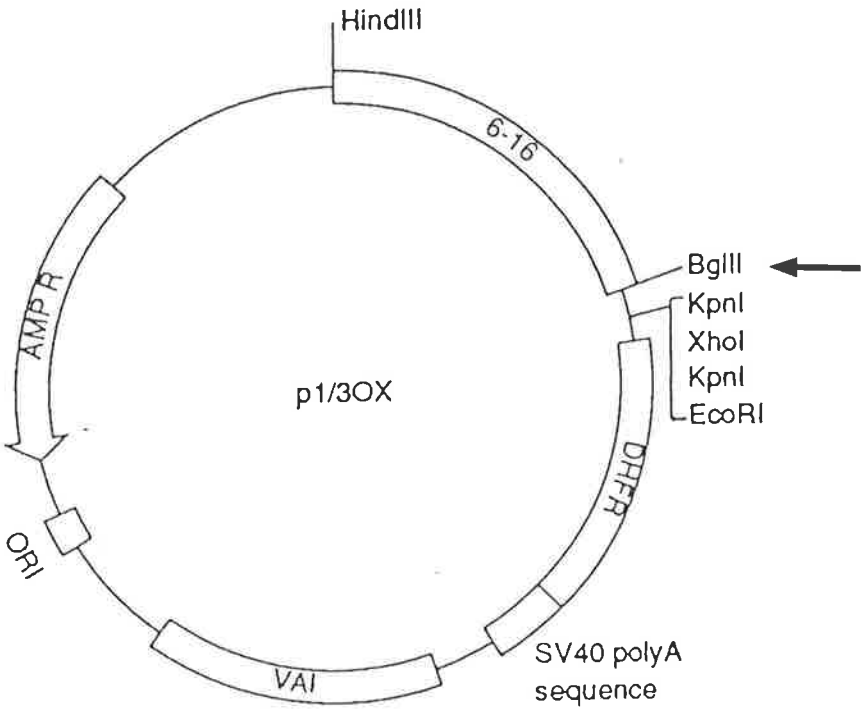
A


H S P Sa
P P
B
N
B B Bg
-603
ISRE
T
EXON 1
INTRON 1
+437

B


H S P Sa X P Hc A X B
P P
B
N B Bg
-603
ISRE
T
+42

C


HindIII
6-16
BglII
KpnI
XhoI
KpnI
EcoRI
p1/3OX
DHFR
SV40 polyA
sequence
VAI
ORI
AMP R

p3OT vectors (Fig. 6.1) to generate the p1Hesx1X and p3Hesx1X constructs. p1Hesx1X and p3Hesx1X stable ES cell lines were generated by co-transfection of MBL-5 ES cells with a 5:1 ratio (20 μg and 4 μg) of expression vector DNA to pPGKNeoR (Smith, 1991). Nine and eight of the p1Hesx1X and p3Hesx1X transfected Neo^R colonies, respectively, were isolated and analysed for integrated plasmid copy number. Approximately 5 μg of genomic DNA derived from each of the Neo^R cell lines was digested with *Pst* I, electrophoresed on a 2% agarose gel and blotted (Fig. 6.2A). This southern blot was probed with a 476 bp *Pst* I fragment common to the p1OT and p3OT vectors derived from the 5' end of the 6-16 promoter fragment (Fig. 6.1), and washed in 2 X SSC / 0.1% SDS at room temperature for 45 minutes followed by 2 X SSC / 0.1% SDS at 68°C for 30 minutes.

A 476 bp band corresponding to integrated p1Hesx1X and p3Hesx1X vector was detected in 1Hesx1X clones 2, 4 and 7 (1Hesx1/2, 1Hesx1/4 and 1Hesx1/7) and in 3Hesx1 clones 1-4 and 7-9 (3Hesx1/1-4, 3Hesx1/7-9; Fig. 6.2B). The detection of larger bands in these tracks was probably due to incomplete digestion of the genomic DNA. The absence of a 476 bp signal in the remaining clones may indicate integration of only the pPGKNeoR vector. Alternatively, insufficient copies of the p1Hesx1X and p3Hesx1X vectors may have been integrated to allow detection. The 476 bp band was of greatest intensity in clones 7 and 9 of the p1Hesx1X and p3Hesx1X clones, respectively. As the 1Hesx1/7 and 3Hesx1/9 tracks did not contain greater quantities of genomic DNA compared with the other clones (Fig. 6.2A), the greater intensity of the 476 bp signal indicated that these clones contained the highest levels of integrated plasmid. Therefore, the 1Hesx1/7 and 3Hesx1/9 stable cell lines were selected for analysis of *Hesx1* induction.

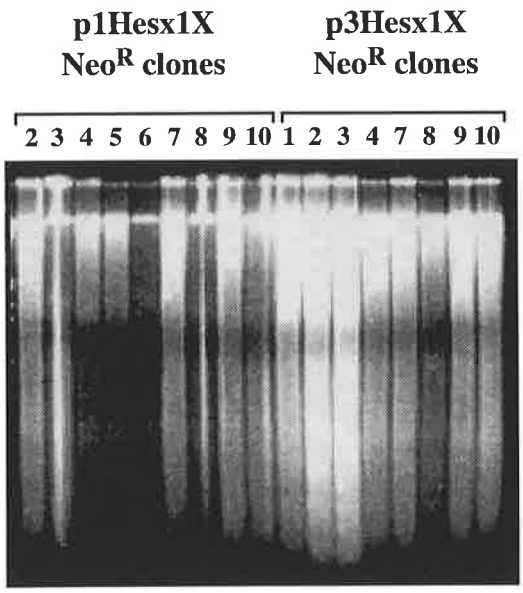
6.2.2. Analysis of *Hesx1* Induction in 1Hesx1/7 and 3Hesx1/9 Stable ES Cell Lines

Interferon induction analysis of stable ES cell transfectants carrying 1OX and 3OX chloramphenicol acetyltransferase (CAT) reporter constructs, indicated that maximum induction levels were achieved by culture in 1000 U/ml interferon (Whyatt *et al.*, 1993). This study also indicated that transgene expression was rapidly induced after exposure to 1000

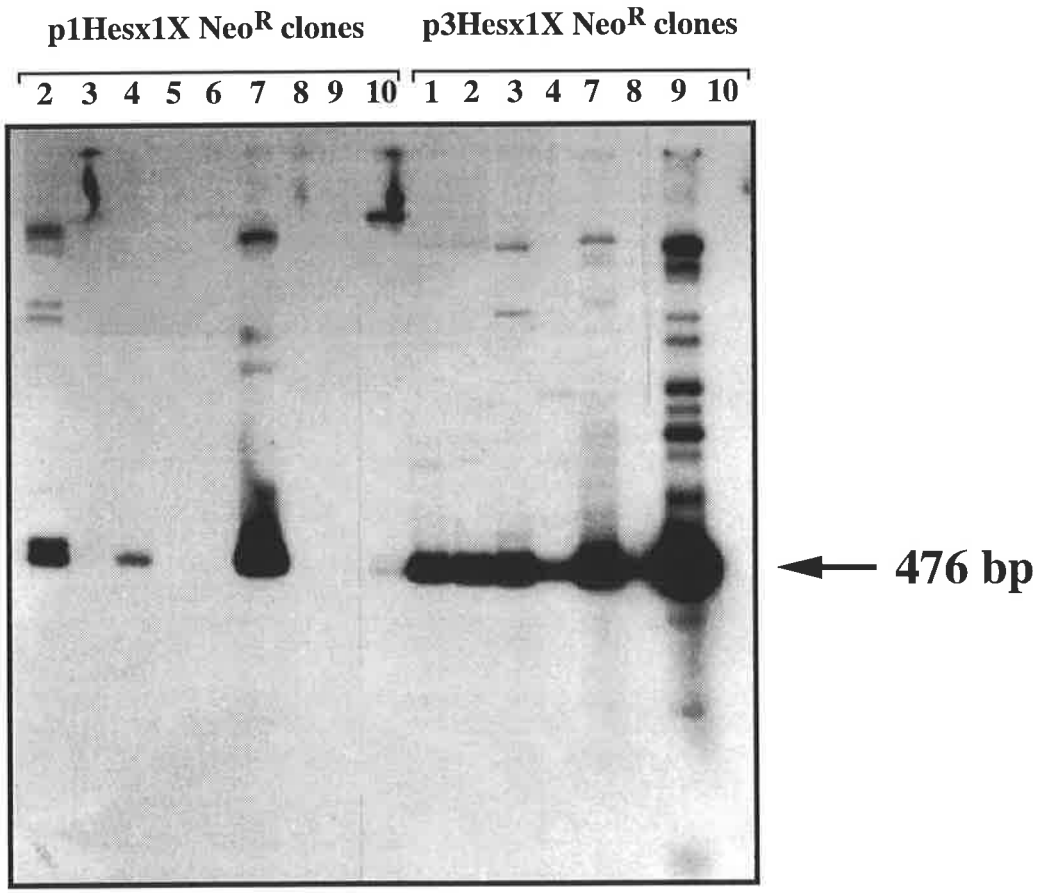
Figure 6.2 Southern blot analysis of integrated copy number in the p1Hesx1X and p3Hesx1X Neo^R clones.

A. Ethidium bromide stained 2% agarose gel containing approximately 5 μg of *Pst* I digested genomic DNA derived from the p1Hesx1X and p3Hesx1X Neo^R clones. **B.** The 2% agarose gel (A) was blotted, and probed with an oligolabelled 476 bp *Pst* I fragment common to the p1OT and p3OT vectors derived from the 5' end of the 6-16 promoter fragment (Fig. 6.1). The southern filter was washed in 2 X SSC / 0.1% SDS at room temperature for 45 minutes followed by 2 X SSC / 0.1% SDS at 68^{o}C for 30 minutes. The 476 bp band corresponding to integrated p1Hesx1X and p3Hesx1X plasmid is indicated by the arrow.

A


p1Hesx1X
Neo^R clones
p3Hesx1X
Neo^R clones
2 3 4 5 6 7 8 9 10 1 2 3 4 7 8 9 10

B


p1Hesx1X Neo^R clones
p3Hesx1X Neo^R clones
2 3 4 5 6 7 8 9 10 1 2 3 4 7 8 9 10
476 bp

U/ml interferon, and that continuous culture with interferon for 24 hours maintained induced expression levels. These parameters were therefore employed for analysis of *Hesx1* induction levels in the 1Hesx1/7 and 3Hesx1/9 stable ES cell lines.

The 1Hesx1/7 and 3Hesx1/9 cell lines were seeded in duplicate at 2 X 10^5 cells/well into 3 cm (diameter) wells in ES cell selection medium. After culture for 24 hours, one half of the cells corresponding to each cell line (one well/cell line) were induced with fresh medium supplemented with 1000 U/ml interferon. The remaining cells received fresh medium without interferon. Cells were cultured for a further 24 hours before harvesting and preparation of cytoplasmic RNA. RNA samples were digested with DNAase 1 to remove contaminating genomic DNA.

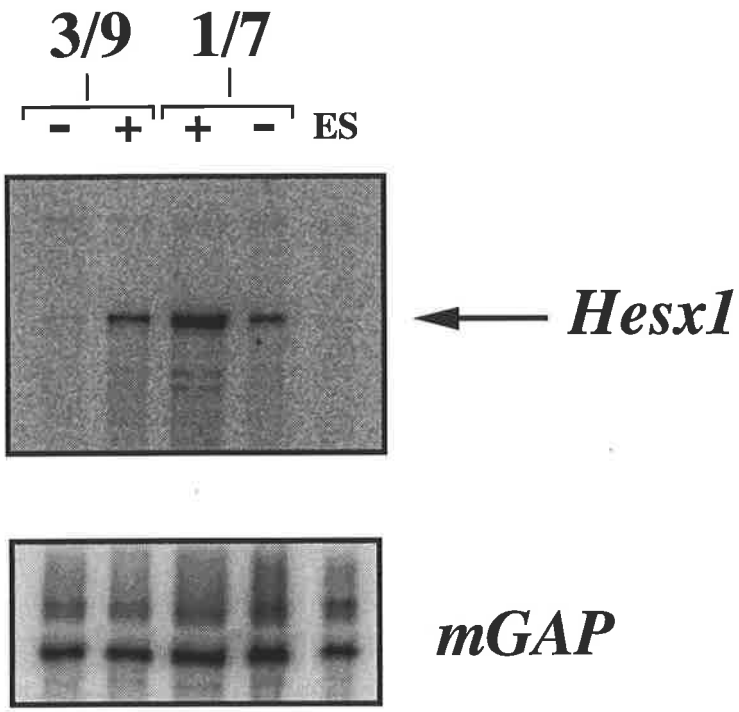
Hesx1 expression levels were assayed by RNAase protection using the 643 bp Hesx1a probe (described in section 4.2.5.) which generated 602 bp and 532 bp bands corresponding to the 1.0 kb and 1.2 kb *Hesx1* transcripts, respectively. Induced *Hesx1* expression derived from the p1Hesx1X and p3Hesx1X constructs also generated a 532 bp signal (as these constructs were derived from the Hesx1b cDNA). 18 µg and 10 µg of RNA derived from the 1Hesx1/7 and 3Hesx1/9 cell lines respectively were assayed. As a positive control, 20 µg of MBL-5 ES cell RNA which had previously been shown to express detectable levels of *Hesx1* (Fig. 4.6) was also included.

Expression of the 1.0 kb and 1.2 kb *Hesx1* transcripts was not detected in the undifferentiated MBL-5 ES cell RNA sample (Fig. 6.3A). This indicated that endogenous *Hesx1* expression was below the level of detection in this experiment. Therefore, *Hesx1* expression in the 1Hesx1/7 and 3Hesx1/9 cells was attributed to expression from the transfected construct. A 532 bp band was detected in the uninduced and induced 1Hesx1/7 and 3Hesx1/9 cells (Fig. 6.3A). Vector derived *Hesx1* expression was detected at a very low level in the uninduced 3Hesx1/9 cell line. However, a relatively high basal expression was detected in the uninduced 1Hesx1/7 line. This higher basal expression level of the p1OT versus the p3OT derived *Hesx1* vectors was consistent with previous analysis of these expression vectors using chloramphenicol acetyltransferase (CAT) reporter constructs

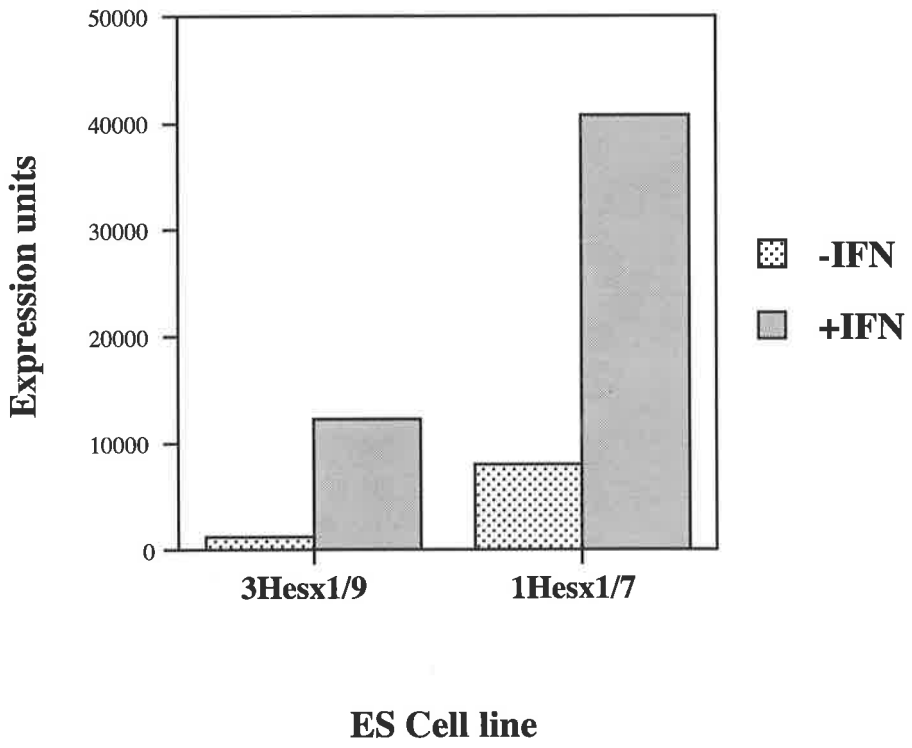
Figure 6.3 Interferon induction analysis of the 1Hesx1/7 and 3Hesx1/9 stable ES cell lines.

A. 1Hesx1/7 and 3Hesx1/9 ES cells cultured in the presence of LIF were induced for 24 hours with 1000 U/ml interferon. 10 μg and 18 μg of RNA derived from the 3Hesx1/9 (3/9) and 1Hesx1/7 (1/7) cell lines, respectively, were assayed for *Hesx1* expression by RNAase protection using the Hesx1a riboprobe (see 4.2.5). 20 μg of MBL-5 ES cell RNA (ES lane) was also assayed. The 532 bp band corresponding to induced *Hesx1* expression is indicated by the arrow. A mGAP loading control was also included (2.3.21). -=uninduced; +=induced. **B.** Graph showing volume integration values of the uninduced and induced *Hesx1* expression in the 1Hesx1/7 and 3Hesx1/9 cell lines.

A


3/9
1/7
- + + - ES
Hesx1
mGAP

B


50000
40000
30000
20000
10000
0
Expression units
3Hesx1/9
1Hesx1/7
ES Cell line
-IFN
+IFN

(Whyatt *et al.*, 1993). Exposure of the 1Hesx1/7 and 3Hesx1/9 cell lines to 1000 U/ml interferon resulted in increased levels of *Hesx1* expression in both of these ES cell lines.

Bands corresponding to uninduced and induced *Hesx1* expression for the 1Hesx1/7 and 3Hesx1/9 cell lines were quantitated by volume integration. The absolute uninduced and induced expression levels for the 1Hesx1/7 and 3Hesx1/9 cell lines were 7796 expression units versus 40734 expression units and 1026 expression units versus 12057 expression units, respectively (Fig. 6.3B). Therefore, the induction ratios for the 1Hesx1/7 and 3Hesx1/9 cell lines were 5.2 fold and 11.7 fold, respectively. The lesser induction ratio of the 3OX derived construct compared with the 1OX derived construct was consistent with previous studies (Whyatt *et al.*, 1993).

As *Hesx1* expression was not detected in the MBL-5 ES cell sample, quantitative comparison between induced and endogenous *Hesx1* expression levels was not possible. However, the detection of induced *Hesx1* expression in the 3Hesx1/9 line indicated that this level of expression was considerably greater than normal endogenous *Hesx1* expression. As the 3Hesx1/9 background (uninduced) *Hesx1* expression was extremely low, this line was selected for *Hesx1* over-expression analyses. The high background level of *Hesx1* expression for the 1Hesx1/7 cells indicated that this line was unsuitable for *Hesx1* induction studies. However, as the background *Hesx1* expression level exceeded normal endogenous expression levels, this line was used for analysis of constitutive *Hesx1* over-expression.

6.3 Investigation of *Hesx1* Over-expression on ES cell Differentiation

6.3.1 Morphology and Differentiation Properties of 1Hesx1/7 and 3Hesx1/9 ES Cell Lines

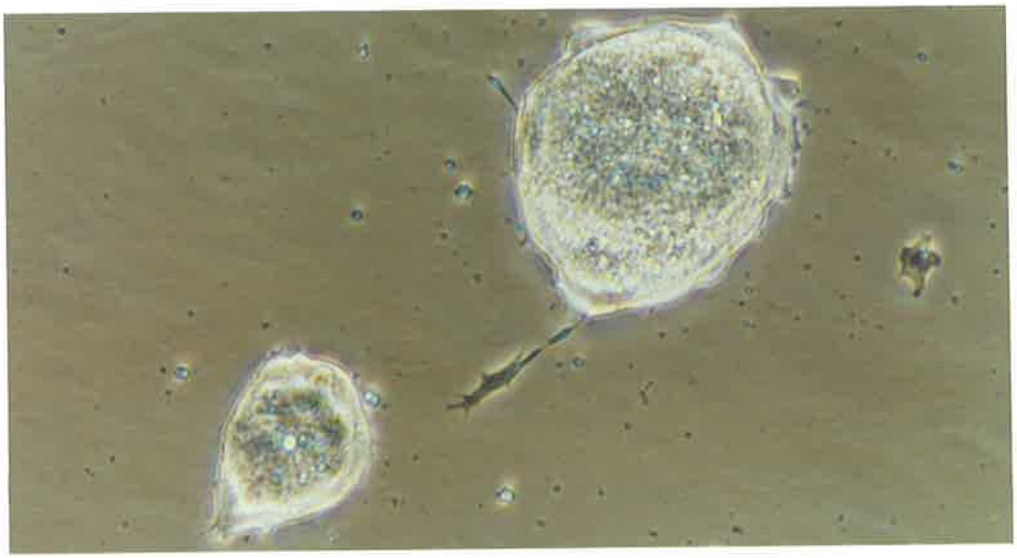
The 1Hesx1/7 and 3Hesx1/9 stable ES cell lines cultured in the presence of LIF, formed dome-shaped ES cell colonies which were indistinct morphologically from the parental MBL-5 ES cell line (Fig. 6.4A,B,C).

ES cells cultured in the presence of LIF give rise to a low level of spontaneously differentiated cells, at the periphery of the dome-shaped ES cell colony. At low frequency,

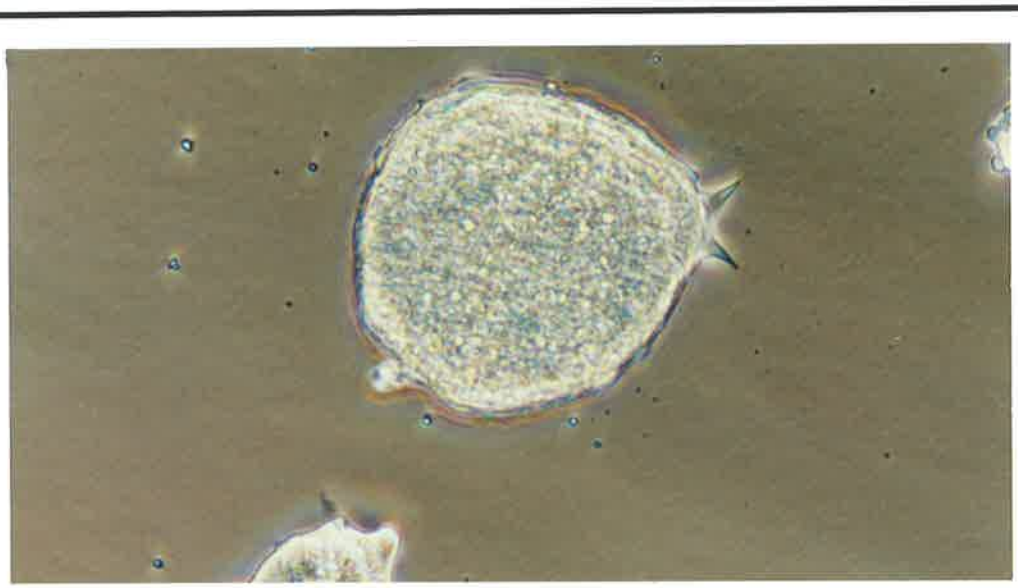
Figure 6.4 Comparison of the colony morphology of MBL-5, 1Hesx1/7 and 3Hesx1/9 ES cell lines.

ES cell colonies of the MBL-5 (A), 1Hesx1/7 (B) and 3Hesx1/9 (C) ES cell lines cultured for three days in the presence of LIF. The cells were photographed under phase contrast optics at 50 X magnification.

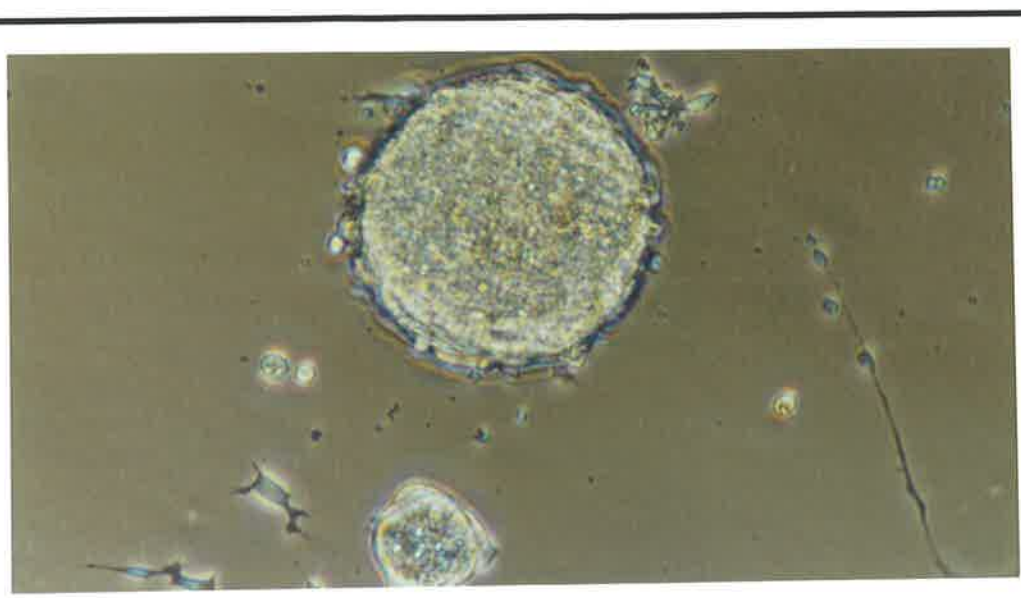
A



B



C



ES cells may also grow as spread colonies, which in my hands appear to be less stable in culture, and readily give rise to peripheral spontaneously differentiated cells of varying morphologies (Fig. 6.5A). The 3Hesx1/9 stable ES cell line was similar to the parental MBL-5 ES cell line in its ability to form spread colonies and to spontaneously differentiate. In contrast, a major difference between the 1Hesx1/7 and MBL-5 cell lines was the reduced propensity of the 1Hesx1/7 cells to generate spontaneously differentiated cells. Although similar proportions of dome-shaped and spread ES cell colonies were observed in 1Hesx1/7 and MBL-5 ES cell cultures, differentiated 1Hesx1/7 ES cells were rarely detected at the periphery of either of these colony types (Fig. 6.4B, Fig 6.5B).

Clonal variation between individual isolates is often observed in the generation of stable pluripotent cell lines in culture (Joy Rathjen, personal communication), although the reason for this phenomenon is not known. However, it is interesting that chromosomal abnormalities have been detected during cytogenetic analysis of independently derived ES cell lines (Robertson and Bradley, 1988). It is possible, therefore, that mutation events which occurred during selection of the 1Hesx1/7 and 3Hesx1/9 cell lines may have resulted in the altered stability of these ES cell lines.

6.3.2 Analysis of *Hesx1* Over-expression on ES Cell Monolayer Differentiation

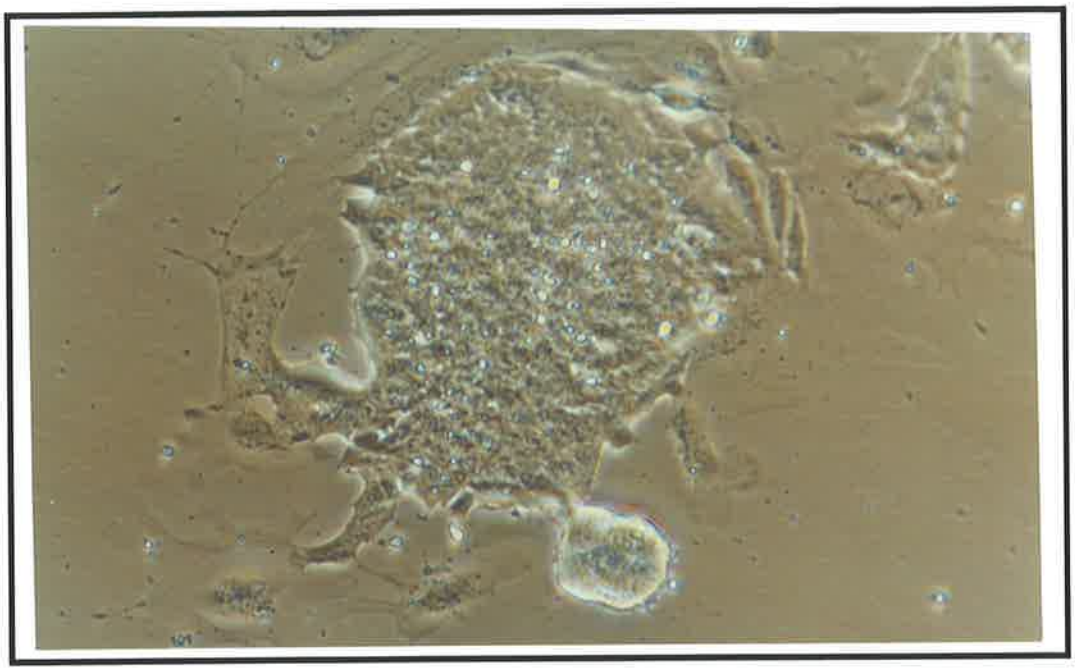
Hesx1 expression data presented in Chapters 3 and 4 indicated that *Hesx1* was expressed by undifferentiated ES cells, and was downregulated during ES cell differentiation. This expression profile suggested that *Hesx1* may play a role in the maintenance of the stem cell phenotype. This possibility was tested directly by comparison of ES cell differentiation between uninduced and induced 3Hesx1/9 cells cultured in a range of LIF concentrations.

Duplicate cultures of 3Hesx1/9 cells were seeded at clonal density into 10 cm dishes. After culture for 48 hours, fresh medium containing 1000 U/ml interferon was added to one of the cultures (induced), and the other culture received fresh medium without interferon (uninduced). After culture for a further 24 hours (to allow maximum levels of *Hesx1*

Figure 6.5 Stability of the MBL-5 and 1Hesx1/7 ES cell lines.

Shown are the spread ES cell colonies and differentiated ES cells which arise spontaneously during culture of the MBL-5 (A) and 1Hesx1/7 (B) cell lines in the presence of LIF. The cells were photographed under phase contrast optics at 50 X magnification.

A



B

induction), the cells were harvested and counted, before seeding in duplicate at 1500 cells/well into ES cell medium containing 1000, 100, 10, 1, 0.1 and 0 U LIF/ml. The ES cell medium for the induced ES cell culture also contained 1000 U/ml interferon. The cells were cultured for a further 5 days before staining for the pluripotent cell marker alkaline phosphatase (Adamson, 1988) to detect the presence of undifferentiated ES cells. Induced and uninduced cells from duplicate cultures of the parental MBL-5 ES cell line were also included as controls.

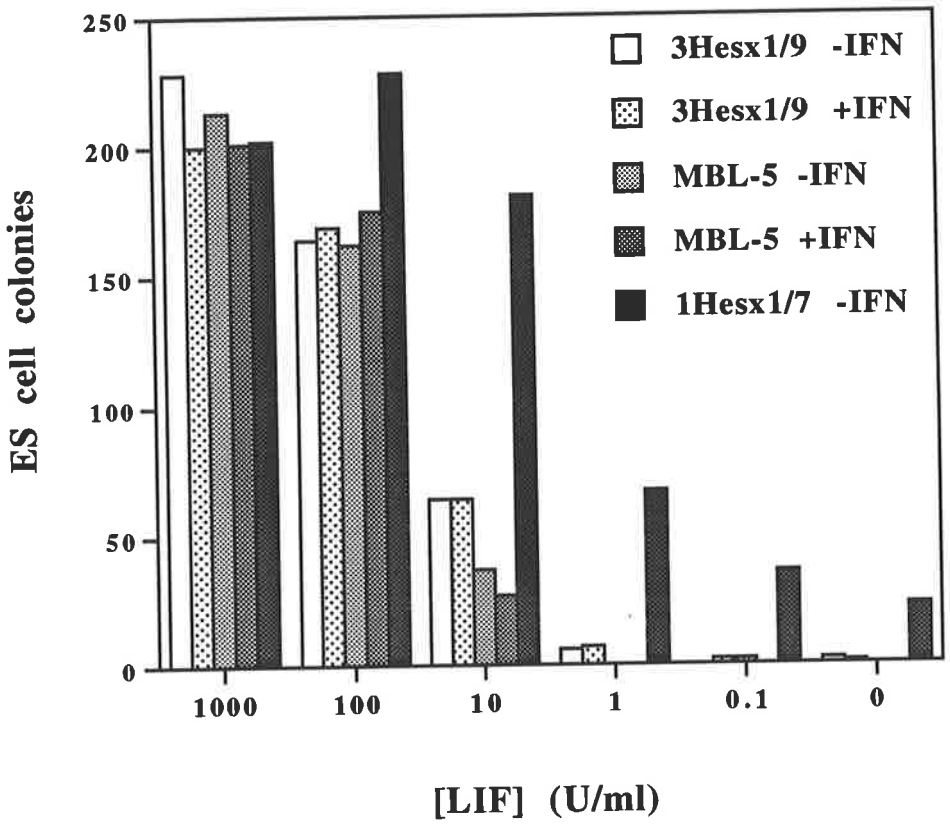
Decreasing numbers of ES cell colonies were observed for both cell lines as the LIF concentration was reduced (Fig. 6.6). The only difference between the cell lines was an approximate two-fold increase in undifferentiated ES cell colonies in the 3Hesx1/9 versus MBL-5 cells at 10 U LIF/ml. This indicated that the 3Hesx1/9 line is more stable at intermediate LIF concentrations than the parental line. However, as this increased stability was detected in both uninduced and induced 3Hesx1/9 cells, it did not appear to be related to *Hesx1* expression levels. No significant difference between the uninduced and induced MBL-5 cells was observed, indicating that exposure to interferon does not alter ES cell stability. This was consistent with the findings of Whyatt *et al.* (Whyatt *et al.*, 1993)

As described in section 6.2.2, the 1Hesx1/7 ES cell line had a high basal (uninduced) *Hesx1* expression level. To investigate the effect of constitutive *Hesx1* over-expression, the LIF dilution assay described above was also carried out in a parallel experiment using the 1Hesx1/7 ES cell line in the absence of interferon.

At 1000 U LIF/ml, similar numbers of undifferentiated ES cell colonies compared with the MBL-5 and 1Hesx1/7 cells were observed (Fig. 6.6). As LIF concentration was reduced from 100 U/ml to 0 U/ml, the number of undifferentiated 1Hesx1/7 ES cells colonies also decreased. This indicated that 1Hesx1/7 cells were multipotent, and required LIF for the maintenance of the undifferentiated state. However at lower LIF concentrations, significantly greater numbers of 1Hesx1/7 ES cell colonies remained compared with 3Hesx1/9 and MBL-5 cells. The increase in undifferentiated ES cell colonies at lower LIF concentrations was consistent with the reduced propensity of 1Hesx1/7 cells to generate spontaneously differentiated cells during routine culture (6.3.1). As the 1Hesx1/7 line had a high basal level

Figure 6.6 Analysis of MBL-5, 1Hesx1/7 and 3Hesx1/9 ES cell monolayer differentiation.

Approximately 1500 uninduced and induced 3Hesx1/9 and MBL-5 ES cells, and uninduced 1Hesx1/7 ES cells, were seeded into ES cell media containing 1000, 100, 10, 1, 0.1 and 0 U/ml LIF. After culture for 5 days, undifferentiated ES cell colonies were detected by alkaline phosphatase staining and counted.


ES cell colonies
250
200
150
100
50
0
1000
100
10
1
0.1
0
[LIF] (U/ml)
3Hesx1/9 -IFN
3Hesx1/9 +IFN
MBL-5 -IFN
MBL-5 +IFN
1Hesx1/7 -IFN

of *Hesx1* expression, one explanation for the increased stability of the 1Hesx1/7 cell line is that it was due to an elevated level of Hesx1 protein. However, as the induced 3Hesx1/9 cell line, which had a similar level of *Hesx1* expression, was not as stable as the (uninduced) 1Hesx1/7 cell line, this possibility seems unlikely. An alternative explanation is that the stability of the 1Hesx1/7 cell line at lower LIF concentrations was an inherent property of this cell line, which reflected clonal variation between individual ES cell isolates.

6.3.3 Investigation of *Hesx1* Over-expression on Embryoid Body Formation and Differentiation

ES cells cultured in suspension form globular structures termed embryoid bodies, which share similarities with the early murine embryo (1.4.3.2). Embryoid bodies have the capacity to generate terminally differentiated cell types which resemble embryonic structures derived from distinct germ-layers, such as beating muscle (mesoderm), blood islets (mesoderm) and neurons (ectoderm) (Doetschmann *et al.*, 1985). Embryoid bodies therefore provide a useful experimental tool to study pluripotent cellular differentiation within a complex "embryo-like" environment. As beating muscle is readily detected in differentiated embryoid body cultures, this tissue provides an assay for terminal cellular differentiation. The following experiment was carried out to investigate the possible effect of *Hesx1* over-expression on embryoid body formation and differentiation.

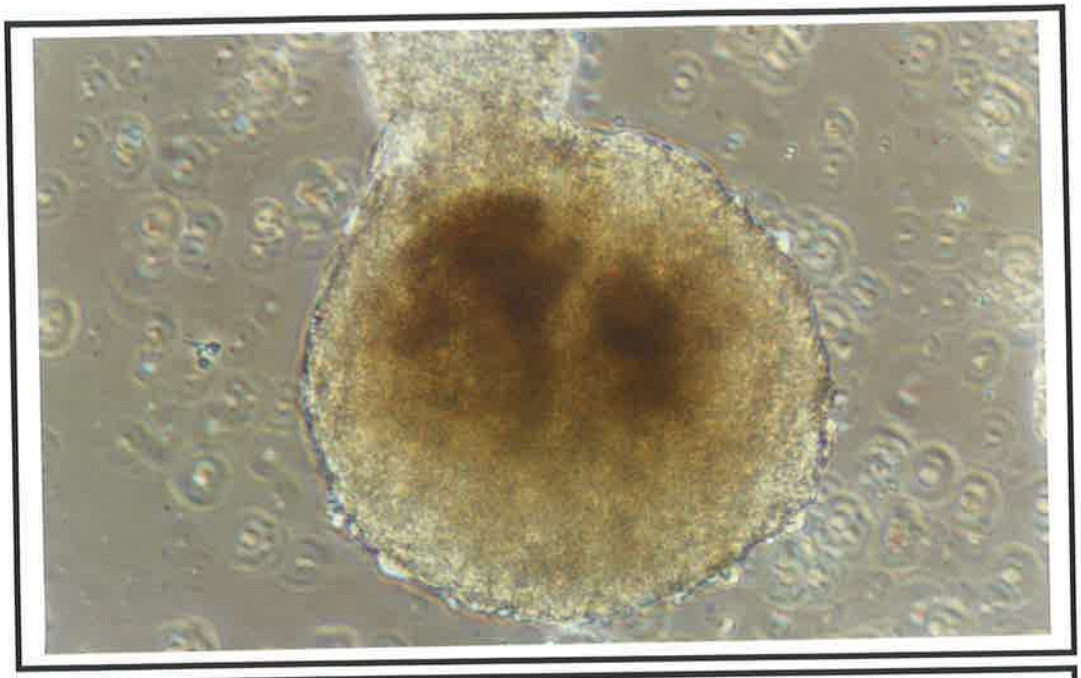
3Hesx1/9 ES cells were induced with 1000 U/ml interferon for 24 hours (as described in 6.2.3) prior to formation of embryoid bodies. Induced cells were maintained in the presence of interferon throughout the experiment. Uninduced 3Hesx1/9 (negative control) and 1Hesx1/7 (constitutive over-expression) cells were also included. Simple embryoid bodies (see 1.4.3.2) were cultured for 5 days before seeding. 48 embryoid bodies from each dish (144 bodies in total) were then seeded into individual wells of a 96 well tray (day 0) and cultured for a further 14 days. Embryoid bodies were inspected daily, or every two days, for the formation of beating muscle.

Simple embryoid bodies cultured for 4 days are shown in Fig. 6.7. No structural difference between the embryoid bodies derived from the induced and uninduced 3Hesx1/9

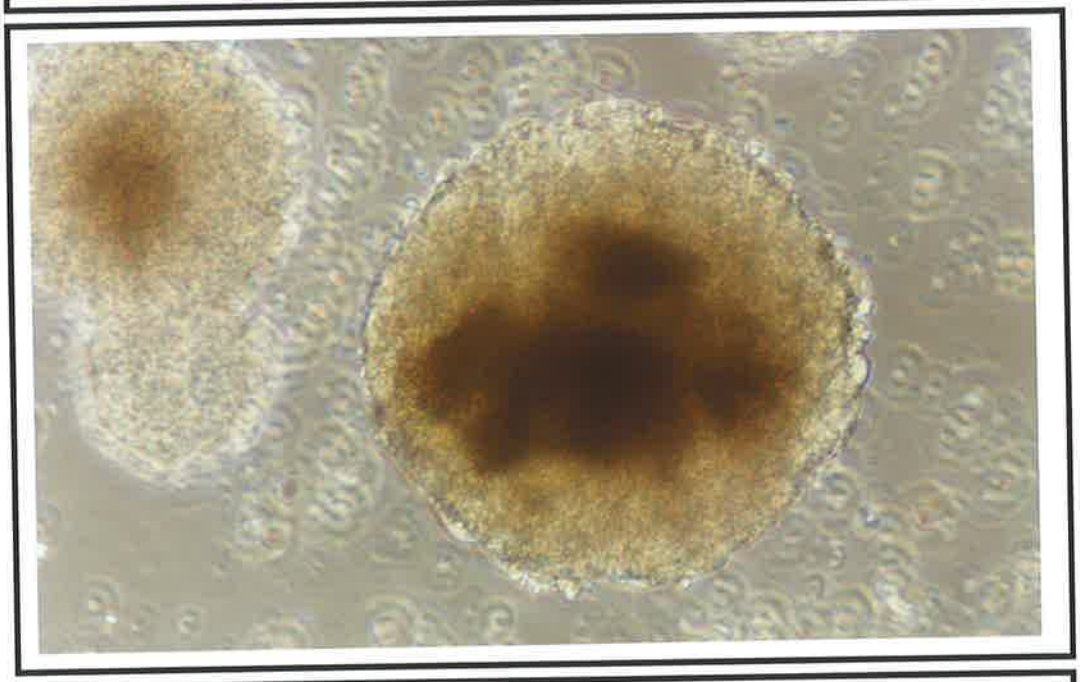
Figure 6.7 3Hesx1/9 and 1Hesx1/7 simple embryoid bodies.

Simple embryoid bodies prepared from uninduced 3Hesx1/9 cells (A), induced 3Hesx1/9 cells (B) and uninduced 1Hesx1/7 cells (C). The embryoid bodies were cultured in suspension for 5 days in the absence of LIF, and photographed under phase contrast at 50 X magnification.

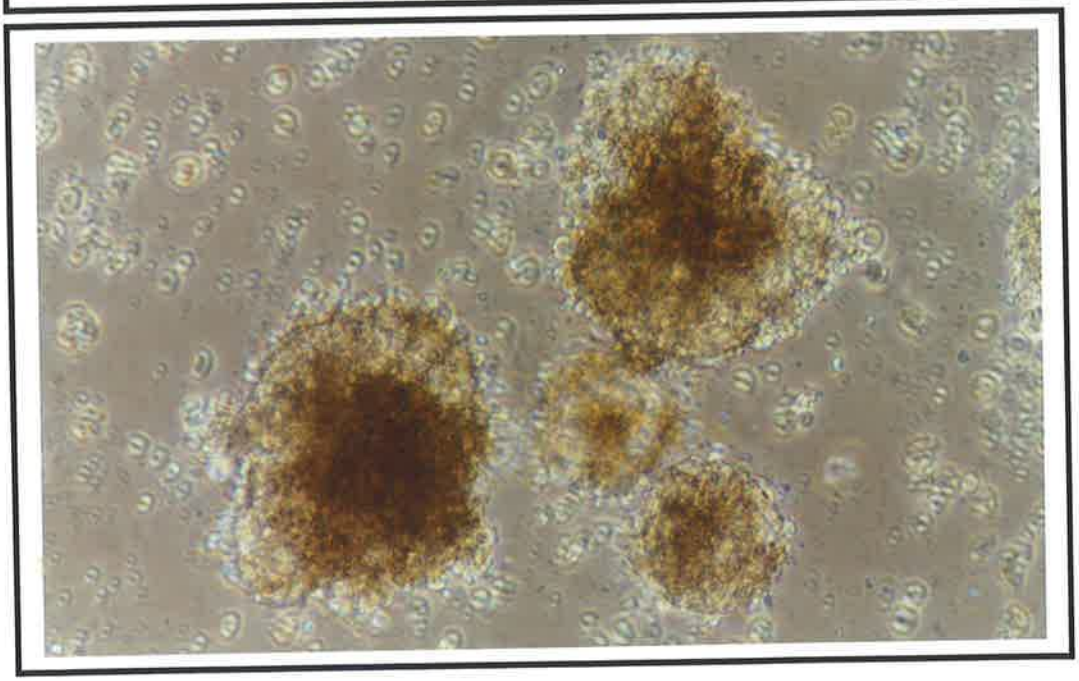
A



B



C



cells were observed (Fig. 6.7A,B). After 4 days in culture, almost all of the 3Hesx1/9 embryoid bodies had developed a discrete outer layer which enclosed a dense core of undifferentiated cells (Robertson, 1987). Embryoid bodies derived from the uninduced 1Hesx1/7 cells were much smaller and less organised (Fig. 6.7C). The high number of dead cells in the medium, coupled with a "ragged" appearance, indicated that the 1Hesx1/7 embryoid bodies were breaking up and releasing dead or dying cells.

The frequency of beating muscle formation is shown in Figure 6.8. No significant difference in embryoid body differentiation between the induced and uninduced 3Hesx1/9 embryoid bodies was observed. Beating muscle was first detected on day 7 in 31% and 25% of the uninduced and induced 3Hesx1/9 embryoid bodies respectively, which is slightly later than has been observed in similar assays (Wang *et al.*, 1992; Lints *et al.*, 1993). After a further day in culture, 79% of the embryoid bodies contained contracting cardiac muscle, and this level (approximately) was maintained over the next 3 days. A range of differentiated cells (Fig. 6.9A,B), including neurons, were also observed around the periphery of the uninduced and induced embryoid bodies indicating that the 3Hesx1/9 cells were multipotent. These results suggest that *Hesx1* over-expression had no effect on embryoid body differentiation.

Seeded 1Hesx1/7 embryoid bodies did not differentiate, and rapidly degenerated into small, dark clumps of dead cells (data not shown). This inability of 1Hesx1/7 embryoid bodies to differentiate, coupled with their poor formation, indicated that the 1Hesx1/7 cell line had a reduced capacity for differentiation. The inability of 1Hesx1/7 embryoid bodies to differentiate was consistent with the enhanced stability of the 1Hesx1/7 cells detected in the LIF dilution assay (6.2.3) and during routine culture (6.3.1).

6.4 Discussion

A preliminary investigation of *Hesx1* function in ES cells was carried out by analysis of *Hesx1* over-expression on ES cell differentiation. Interferon inducible stable ES cell lines were generated using the p1OX and p3OX expression vectors which were capable of 5.2 (1Hesx1/7) and 11.7 (3Hesx1/9) fold induction of *Hesx1* expression. As the uninduced

Figure 6.8 Comparison of beating muscle formation by uninduced and induced 3Hesx1/9 embryoid bodies.

48 simple embryoid bodies prepared from uninduced and induced 3Hesx1/9 cells were cultured for 5 days before seeding into individual wells of a 96 well tray (day 0) Embryoid bodies were cultured for a further 14 days and inspected daily, or every two days, for the presence of beating muscle. The vertical axis shows the percentage of cultures which contained beating muscle.

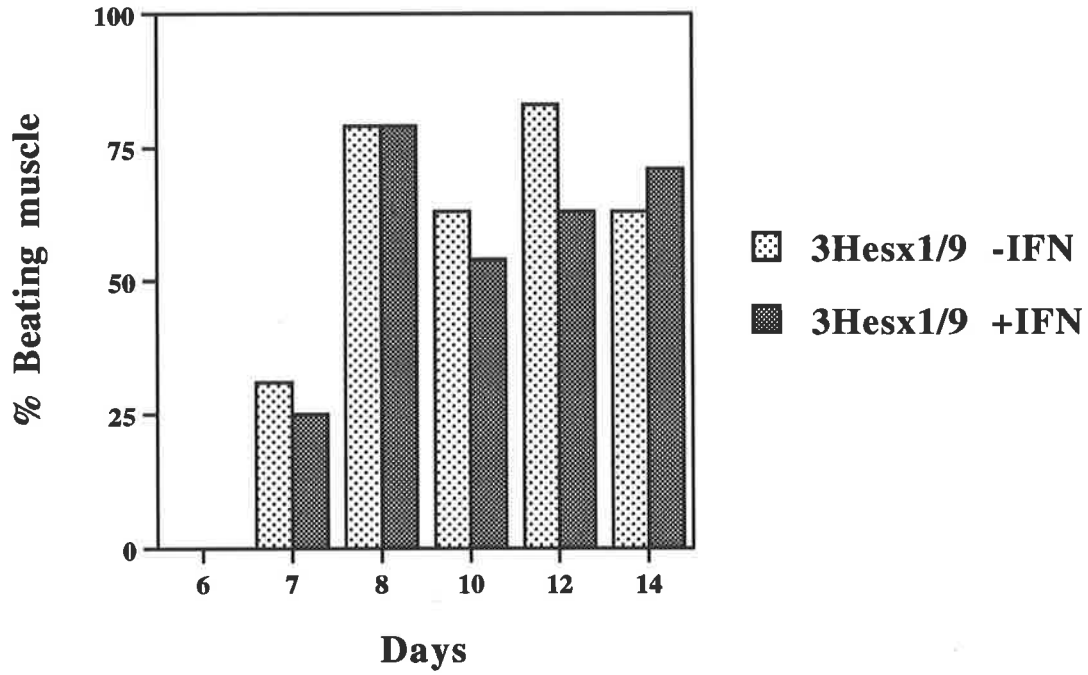
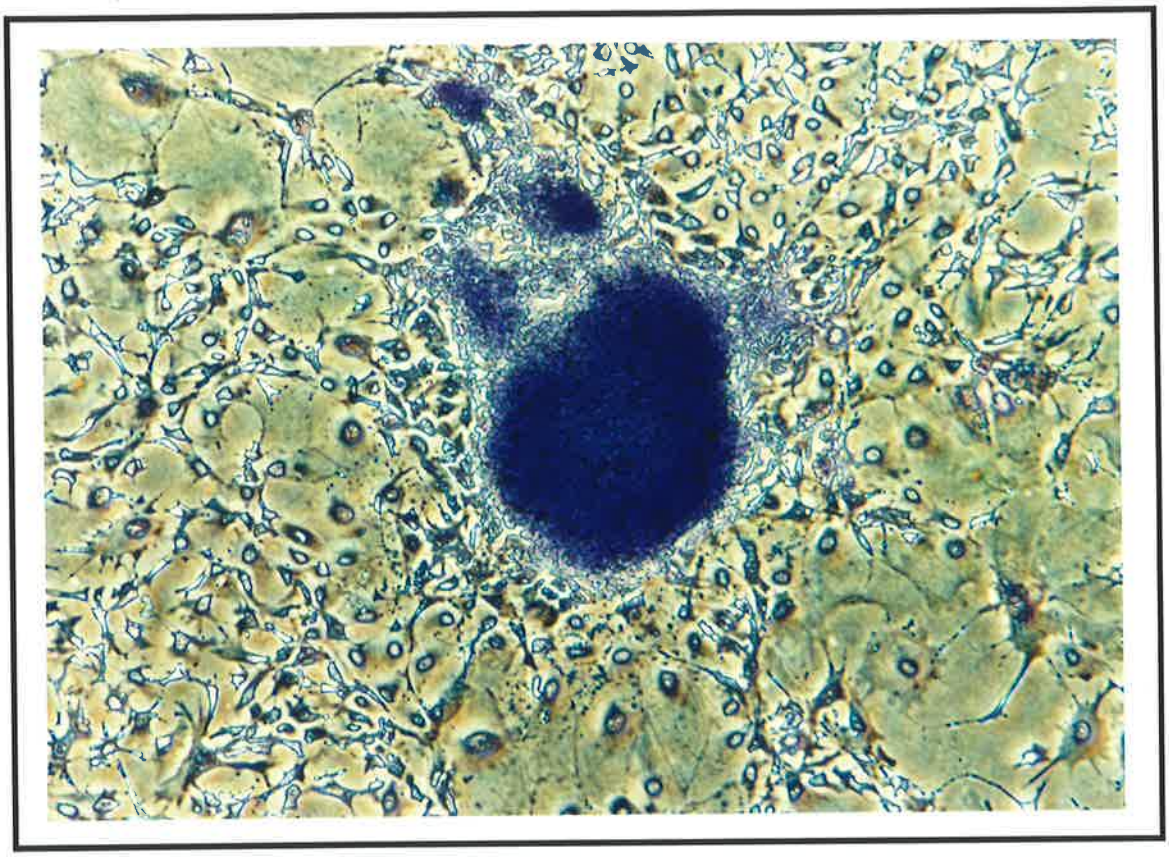

100
75
50
25
0
% Beating muscle
6
7
8
10
12
14
Days
3Hesx1/9 -IFN
3Hesx1/9 +IFN

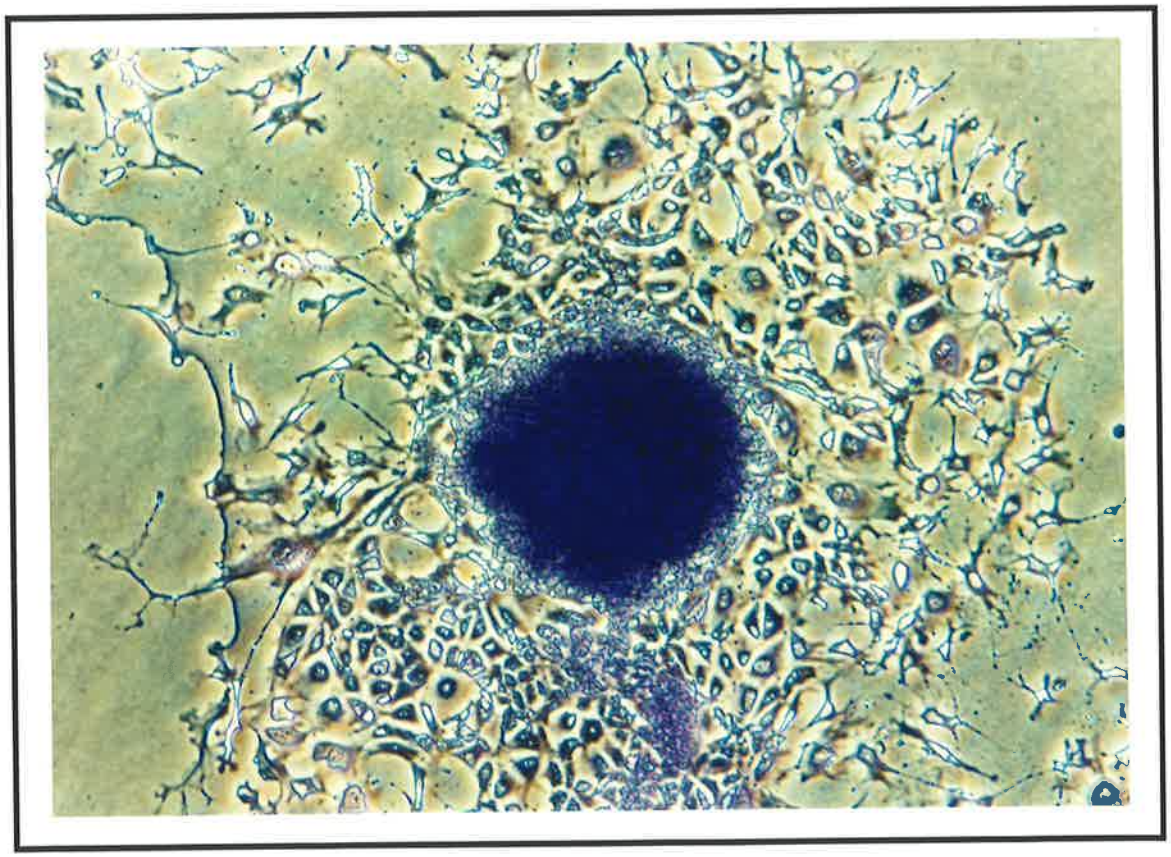
Figure 6.9 Differentiated cell types generated by uninduced and induced 3Hesx1/9 seeded embryoid bodies.

Uninduced and induced 3Hesx1/9 simple embryoid bodies were seeded and cultured for 14 days in the absence of LIF. A variety of differentiated cell types were detected about the periphery of the uninduced (A) and induced (B) 3Hesx1/9 seeded embryoid bodies. The cells were stained with Giemsa and photographed under phase contrast optics at 50 X magnification.

A



B



Hesx1 expression in the 3Hesx1/9 cell line was barely detectable (Fig. 6.3A), further analysis of uninduced and induced *Hesx1* expression using longer exposure times may be required to achieve a more accurate estimate of the induction ratio. The 1Hesx1/7 and 3Hesx1/9 induction levels were less than has previously been observed with p1OT and p3OT chloramphenicol acetyltransferase (CAT) reporter constructs (Whyatt *et al.*, 1993), which may indicate some inherent instability in the (induced) *Hesx1* transcript.

The Hesx1a cDNA riboprobe used in the induction analysis did not detect a 602 bp band corresponding to the 1.0 kb *Hesx1* transcript in either the uninduced or induced stable ES cell lines. Therefore, no correlation between the levels of endogenous and induced *Hesx1* expression was observed. Although it has not been established that induction of the *Hesx1* transcript resulted in elevated levels of *Hesx1* protein (see below), the inability to detect the 1.0 kb transcript in the induced stable ES cell lines suggested that *Hesx1*, unlike many *Drosophila* homeobox genes, is not autoregulatory (Hayashi and Scott, 1990).

Comparison between uninduced and induced 3Hesx1/9 cells in the maintenance of ES cell colonies in decreasing concentrations of LIF (6.3.2), and in embryoid body formation and differentiation (6.3.3) did not reveal any significant differences. The failure to detect a difference between uninduced and induced 3Hesx1/9 cells may be due to at least two factors. Firstly, it may be that insufficient levels of (induced) *Hesx1* expression were present to result in a biological effect. This explanation would be consistent with a study of *Hox-c6* (*Hox 3.3*) expression in mouse breast cancer cells in which cell surface changes were only detected above a threshold limit of expression (Shimeld and Sharpe, 1992). Further investigation should include the isolation of additional stable ES cell lines capable of greater levels of induced *Hesx1* expression. A second possible explanation is that *Hesx1* RNA levels did not correlate with *Hesx1* protein levels. Until an *Hesx1* antibody is available, this possibility remains unresolved.

Functional analysis of *Hesx1* by over-expression in ES cells failed to detect a correlation between *Hesx1* expression and ES cell differentiation. Therefore, the possibility that *Hesx1* expression in ES cells is artifactual, and does not reflect a developmental function for this gene cannot be excluded. Nevertheless, regardless of the functional significance of

Hesx1 expression in ES cells, the downregulation of *Hesx1* expression during ES cell differentiation indicates that *Hesx1* provides a useful ES cell specific gene marker.

The 1Hesx1/7 ES cell line, which had a high level of (uninduced) basal *Hesx1* expression, displayed an reduced capacity to differentiate in lowered LIF concentrations than the parental MBL-5 cell line. Furthermore, embryoid bodies derived from (uninduced) 1Hesx1/7 cells were poorly formed and failed to differentiate when seeded. It is possible that this increased stability is due to the high level of basal *Hesx1* expression. However, this seems unlikely as comparable levels of induced *Hesx1* expression were detected in the 3Hesx1/9 cell line. It seems more likely that the 1Hesx1/7 phenotype is the result of clonal variation which is often observed in the isolation of stable pluripotent stem cell lines (Joy Rathjen, University of Adelaide, personal communication). This variation may also account for the slightly increased stability of the 3Hesx1/9 cells detected in the LIF dilution assay, and the later formation of beating muscle compared with previous studies (Wang *et al.*, 1992; Lints *et al.*, 1993).

From the preliminary analysis described in this chapter, it is still unclear whether there is a correlation between *Hesx1* over-expression and ES cell maintenance or differentiation. It is possible that *Hesx1* over-expression induced subtle changes in cell morphology which were not detected by microscopic analysis. This possibility may bear further investigation using the sensitive aqueous two-phase partition technique (Walter *et al.*, 1985), which has been used to detect subtle changes in cell surface molecules resulting from *Hox-c3* (*Hox 3.3*) over-expression (Shimeld and Sharpe, 1992). This approach may be particularly useful as homeobox genes have been shown to regulate expression cell surface protein genes such as *Ncam* (neural cellular adhesion molecule; Valarche *et al.*, 1993) and *Connectin* (Gould and White, 1992). In addition, possible alterations in gene expression resulting from *Hesx1* over-expression were not investigated. It should prove possible to explore this possibility using the sensitive PCR differential display technique (Liang and Pardee, 1992; Zimmermann and Schultz, 1994), which would also have the potential to identify *Hesx1* target genes. Finally, it may be useful to generate inducible *Hesx1* antisense ES cell lines for the investigation of ES cell behaviour in the absence of *Hesx1*.

CHAPTER 7: GENERAL DISCUSSION

CHAPTER 7: GENERAL DISCUSSION

7.1 General Discussion

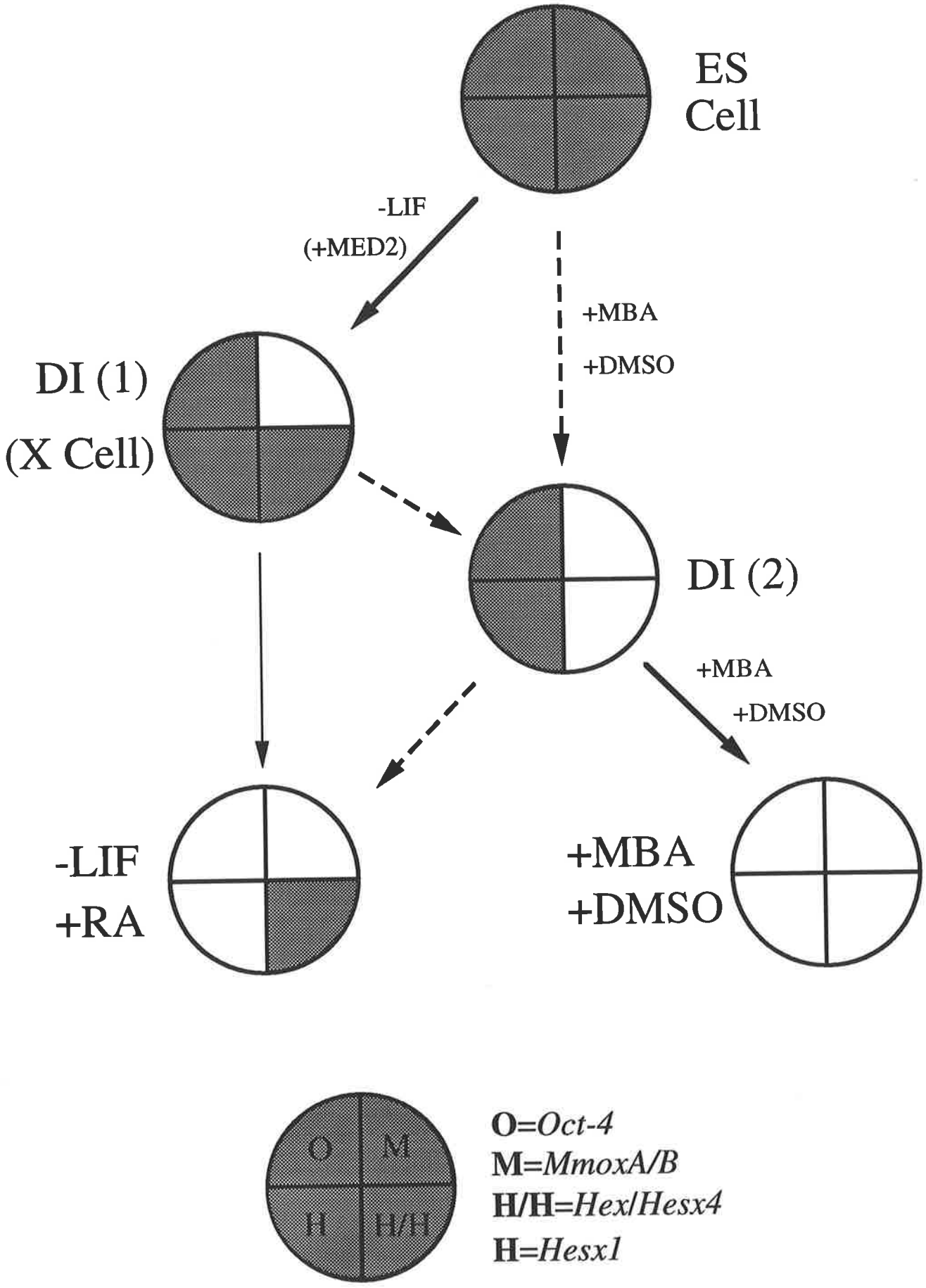
7.1.1 Analysis of Homeobox Gene Expression during ES Cell Differentiation

Extensive analysis of homeobox gene expression and function during the middle and late stages of murine embryogenesis have demonstrated that this class of transcription factors plays a significant developmental role in pattern formation and in the specification of cellular and regional identity (Chapter 1). However, as most studies have focussed on latter stages of development, expression of homeobox genes during early murine ontogeny was largely unknown. The work presented in this thesis was aimed at providing insights into the developmental role of homeobox genes expressed by pluripotent cells during early development.

PCR screening of ES cell cDNA identified thirteen homeobox gene sequences, which included three novel sequences named *Hesx1*, *2* and *4*. Several of the homeobox genes which were isolated, including *Hesx1*, *Hesx4*, *Hex*, *MMoxA* and/or *MMoxB* were subsequently shown to be expressed by ES cells. This demonstrated for the first time that ES cells express a range of homeobox genes, and supports the predicted developmental role of these genes during early development. The homeobox genes which were expressed by ES cells encoded relatively diverged homeodomain sequences which did not belong to established homeodomain sequence classes. The expression of multiple novel diverged homeodomain sequences may reflect the unique biological potential of ES cells, and suggests that pluripotent cells may control developmental decisions such differentiation versus proliferation via a distinct repertoire of homeobox genes compared with the differentiated cells of latter embryogenesis. Furthermore, as genes which encode highly related homeodomain sequences have been shown to be expressed in complimentary or overlapping domains during embryogenesis (Hunt and Krumlauf, 1992; Simeone *et al.*, 1992; Davidson *et al.*, 1988), this diversity may indicate the existence of novel homeobox

Figure 7.1 Model of alternative ES cell differentiation pathways which occur during chemical and spontaneous induction.

The model of ES cell differentiation is based on comparison between the expression of *Oct-4* and *MmoxA/B*, *Hex*, *Hesx4* and *Hesx1* in uninduced and induced ES cells (Chapter 3). Shading indicates the expression of a particular homeobox gene. The arrows indicate discrete differentiation events, and the culture conditions which induce a particular differentiation event are shown next to the arrow. Bold solid arrows indicate differentiation steps for which there was direct evidence. Dashed arrows indicate additional differentiation pathways which may also occur. The greater likelihood that the -LIF / + RA differentiated cells are generated from the DI (1) versus the DI (2) intermediate is indicated by the thin solid arrow between DI (1) and the -LIF / + RA differentiated cells. DI (1)-Differentiation Intermediate 1; DI (2)-Differentiation Intermediate 2.


ES
Cell
-LIF
(+MED2)
+MBA
+DMSO
DI (1)
(X Cell)
DI (2)
+MBA
+DMSO
-LIF
+RA
+MBA
+DMSO
O
M
H
H/H
O=Oct-4
M=MmoxA/B
H/H=Hex/Hesx4
H=Hesx1

gene classes which have restricted expression in pluripotent cells during early development. However, as genomic southern blot analysis using the *Hesx1* partial homeobox sequence detected only a single *Hesx1* specific signal, it is unlikely that additional homeobox genes which belong to the *Hesx1*/XANF-1 sequence class are expressed during early embryogenesis.

Comparison of homeobox gene expression in induced and uninduced ES cells with expression of the pluripotent cell marker *Oct-4*, allowed classification of the homeobox genes expressed by ES cells into distinct groups. The correlation between *Hesx1* and *Oct-4* expression in the spontaneously- and chemically-induced ES cells indicated that *Hesx1* was downregulated during ES cell differentiation, and that expression was restricted to the residual pluripotent (*Oct-4*$^+$) cells. The lack of *MmoxA/B* expression in the pluripotent (*Oct-4*$^+$) X-1 and X+3 cells, and in the residual pluripotent cells after spontaneous or chemical induction with MBA or DMSO, indicated the existence of two distinct groups of *Oct-4*$^+$ pluripotent cells which could be discriminated on the basis of their *MmoxA/B* expression. Similarly, the greater relative downregulation of *Hex* and *Hesx4* versus *Oct-4* in the MBA and DMSO induced ES cell cultures indicated the presence of pluripotent cells with the phenotype *Hex*$^-$/*Hesx4*$^-$/*Oct-4*$^+$. However, *Hex* and *Hesx4* differed from *MmoxA/B* by their expression in the spontaneously- and RA-induced cells. These results suggest that distinct populations of pluripotent cells, as defined by *Oct-4* expression, may be identified by their homeobox gene expression profile. Furthermore, the differing expression patterns of these homeobox genes in induced ES cells suggests that these genes have distinct developmental roles during pluripotent stem cell differentiation *in vivo*.

A model of ES cell differentiation based on the homeobox gene expression data is proposed (Fig. 7.1). ES cells cultured in the presence of LIF express *MmoxA/B*, *Hex*, *Hesx4*, *Hesx1* and *Oct-4*. Two distinct Differentiation Intermediates, DI (1) and DI (2), are proposed to occur during ES cell differentiation.

DI (1): In comparison with *Oct-4*, high levels of *Hex* and *Hesx4* expression, and low levels of *MmoxA/B* expression, were detected in spontaneously (-LIF) induced ES cells. This expression pattern implies the existence of the DI (1) differentiation intermediate (*Oct-*

4^{+}*/MmoxA/B*$^{-}$*/Hex*$^{+}$*/Hesx4*$^{+}$*/Hesx1*$^{+}$), which differs from ES cells by the loss of *MmoxA/B* expression. This Differentiation Intermediate may be equivalent to X-1 and X+3 cells (X cells) which also express greatly reduced levels of *MmoxA/B*.

DI (2): An alternative differentiation event, induced by exposure of ES cells to MBA or DMSO, resulted in the loss of *MmoxA/B*, *Hex* and *Hesx4* expression and the formation of the DI (2) Differentiation Intermediate. It is unclear whether ES cells differentiate directly into the DI (2) intermediate, or whether formation of the DI (2) intermediate is preceded by generation of the DI (1) intermediate (this latter possibility is indicated by the dashed arrow between DI (1) and DI (2)). MBA- and DMSO-induction of the DI (2) intermediate generates the +MBA and +DMSO differentiated cell types, with a concomitant loss of *Oct-4* and *Hesx1* expression. It is also formally possible that spontaneous or RA-induction of the DI (2) intermediate may generate the -LIF and +RA differentiated cell types. However it seems more likely that the -LIF and +RA differentiated cell types are generated directly from the DI (1) intermediate, as this pathway does not require downregulation and re-expression of *Hex* and *Hesx4* (indicated by the thin line between DI (1) and the (-LIF, +RA) differentiated cell type).

Each stage in the ES cell differentiation pathway is characterized by a unique homeobox gene expression profile. Therefore, the homeobox genes which have been isolated as part of this work may provide gene markers for the identification of distinct differentiation intermediates during ES cell differentiation. This may be investigated by simultaneous direct visualization of cellular homeobox gene expression in chemically and spontaneously induced ES cell monolayer cultures by *in situ* hybridization analysis. For example, direct comparison of *MmoxA/B*, *Hex*, *Hesx4* and *Oct-4* expression may determine whether the DI (1) intermediate is generated as a precursor to the DI (2) intermediate during MBA- and DMSO-induced ES cell differentiation. This approach may, therefore, also help to resolve the alternative differentiation pathways described in the model.

The restricted expression of *MmoxA/B*, *Hex*, *Hesx4* and *Hesx1* to different differentiation stages, coupled with the ability of homeobox genes to specify cellular identity, suggest that these genes may play distinct developmental roles in the coordination

of pluripotent cell differentiation during early development. More specifically, the restricted expression of *MmoxA/B* to undifferentiated ES cells suggests that *MMoxA* and/or *MMoxB* may play some part in maintaining pluripotent cells in the undifferentiated state. Similarly, the persistence or re-expression of *Hex* and *Hesx4* in RA induced ES cells, coupled with the absence of *Hex* and *Hesx4* expression in the (MBA and DMSO induced) Differentiation Intermediate, suggest that these genes may play a role in the commitment of ES cells towards specific developmental pathways.

In pre-gastrulation embryos, *Oct-4* is expressed by the inner cell mass and the primitive ectoderm from which the entire embryo is derived, and is downregulated during differentiation of these tissues (Rosner *et al.*, 1990; Yeom *et al.*, 1991; Scholer *et al.*, 1990b). The restricted expression of *MMoxA* and/or *MMoxB*, *Hex* and *Hesx4* to a subset of the pluripotent cell population *in vitro*, suggests that these genes may be expressed by only a proportion of the cells which constitute these tissues. Fate mapping studies have suggested that the primitive ectoderm has region specific variation in developmental potential (Beddington, 1982). However, restricted gene expression in specific portions of this tissue has not been detected. The expression studies presented in this thesis suggest that *MmoxA/B*, *Hex* and *Hesx4* may exhibit this expression pattern. Alternatively, these genes may be expressed in the ICM cells and downregulated during primitive ectoderm formation. It is critical, therefore, to map the expression of *MMoxA*, *MMoxB*, *Hex* and *Hesx4* in the early embryo to determine whether expression of these genes is restricted to a sub-population of the pluripotent cells.

Hesx1, *Hex*, *Hesx4* and *MmoxA/B* belong to a relatively small group of homeobox genes which are known to be downregulated during ES cell differentiation. Apart from the functional significance of this expression pattern, these genes could provide ES cell specific markers which will facilitate more rigourous definition of pluripotent and differentiated cell types *in vitro* and *in vivo*. It is expected that *in situ* analysis of *Hesx1*, *Hex*, *Hesx4* and *MmoxA/B* expression will allow direct correlation between cell types and differentiation pathways *in vitro* and *in vivo*. Analysis of marker gene expression *in vitro* and *in vivo* will therefore facilitate the interpretation of pluripotent cell differentiation *in vitro* within an

embryonic context. As the pluripotent cell population from which ES cells are derived has not been rigourously defined, *in situ* analysis of *Hesx1*, *Hex*, *Hesx4* and *MmoxA/B* expression in the early embryo should also provide some insight into the cellular origin of ES cells. From a commercial and agricultural perspective, ES cell specific markers would assist in the isolation of ES cells from domestic and livestock species, and thereby facilitate the genetic manipulation of these species.

7.1.2 Characterization of the Novel Homeobox Gene, *Hesx1*

Hesx1 was the first of the novel homeobox genes to be isolated, and was selected for further analysis on the basis of its restricted expression to pluripotent cells and novel (diverged) homeodomain sequence. cDNA clones corresponding to the 1.0 kb and 1.2 kb *Hesx1* transcripts, and a *Hesx1* genomic clone, were isolated. The difference between the *Hesx1* transcripts was shown to arise from differential utilization of a poor consensus splice donor site immediately downstream of the termination codon. As these transcripts encode identical open reading frames, a distinct biological role for the alternative *Hesx1* transcripts was not immediately apparent. However, the 1.2 kb transcript was expressed by ES cells, but was not detected in tissue specific RNA samples derived from 14.5 days p.c. and 16 days p.c. embryos. This expression pattern suggests that the alternative splicing event which generated the 1.2 kb transcript may be developmentally regulated, and may reflect specific function for this transcript in pluripotent cells.

Comparison between the *Hesx1* and XANF-1 (Zaraisky *et al.*, 1992) open reading frames revealed 47% homology across the entire sequence. This conservation, coupled with the identification of a common termination site, indicated that these genes were derived from a common ancestor. This degree of homology, however, is probably not sufficient to describe these as sequence homologs. The identity between the open reading frames is restricted largely to the homeodomain and the amino acid residues immediately upstream and downstream of the homeodomain. The conservation of these flanking sequences points to a functional role for these residues, possibly in homeodomain/DNA interaction.

In addition to the predicted expression of *Hesx1* in pluripotent cells during early development, the restricted expression of *Hesx1* to embryonic liver at 14.5 days p.c. and 16 days p.c. supports a role for *Hesx1* in foetal haematopoiesis. It is interesting that *Hex*, which is expressed by pluripotent cells, is also expressed by early myeloid and lymphoid progenitor cell lines (Bedford *et al.*, 1993). The expression of homeobox genes at multiple stages of development and/or stages of cellular differentiation has been demonstrated for a variety of homeobox genes (Hunt and Krumlauf, 1992). This indicates that the ability of individual homeobox genes to regulate target gene expression may be used a number times during embryogenesis within different cellular contexts. The expression data indicate that this may also apply to *Hesx1* and *Hex*, and supports a dual (and perhaps related) role for these genes in pluripotent and haematopoietic progenitor cells.

During the course of this work, *Hesx1* sequence was isolated from a variety of mouse strains, including 129/Sv, Balb/c, CBA and B6C3Fe-a/a-Wc. Several distinct *Hesx1* alleles were detected. All the *Hesx1* alleles encoded identical homeodomain sequences, although single silent base changes in the B6C3Fe-a/a-Wc homeobox sequence were detected. Of particular interest were the CBA-1 and CBA-2 *Hesx1* alleles which encoded an R to K substitution at positions 176 and an R to K and a K to R at positions 176 and 177 respectively. An additional non-conservative substitution (asparagine to aspartate) at position 51 was detected in the B6C3Fe-a/a-Wc strain. Variation at these positions (particularly position 51) in the *Hesx1* protein sequence point to a lesser role for these amino acid residues in *Hesx1* structure and/or function. This is supported by the lack of conservation of these residues in the XANF-1 open reading frame.

7.2 Future Work

The existence of distinct pluripotent differentiation intermediates was first suggested by comparison of homeobox gene and *Oct-4* expression using northern blot analysis. This was supported by the subsequent direct analysis of cellular gene expression using *in situ* hybridization. However, identification of the proposed differentiation intermediates requires direct comparison of gene expression in the individual cell. Toward this end it is essential

that a double staining assay for simultaneous visualization of gene expression in ES cell monolayers is developed. This may prove to be possible through the use of radioactive or fluorescent labelling techniques.

The restricted expression of *MmoxA/B* to a subset of pluripotent cells warrants further investigation of these closely related genes. The isolation of *MMoxA* and/or *MMoxB* cDNA clones will provide gene specific probes to allow direct analysis of the relative contribution of each gene to the *MmoxA/B* signal, and provide a basis for the investigation of gene function.

Correlation between the expression studies carried out *in vitro*, and the expression of these genes during early development, requires the investigation of *Hesx1*, *Hesx4*, *Hex*, *MMoxA* and *MMoxB* expression in the early mouse embryo. This should provide further insight into the developmental role of these genes during pluripotent cell differentiation. *In situ* analysis of *MMoxA* and/or *MMoxB* may be particularly useful as these genes should aid the identification of the pluripotent cell population(s) from which ES cells are derived. The proposed role of *Hesx1* in haematopoiesis may also be established by *in situ* hybridization analysis of *Hesx1* expression in foetal liver. In addition, this may be investigated by analysis of *Hesx1* expression in a range of cell lines corresponding to different stages of haematopoiesis.

Finally, one approach which has become fundamental to the analysis of gene function *in vivo* is to generate null mutant mice by gene targeting. This strategy has been instrumental in the dissection of homeobox gene function, particularly for the *Hox* genes (1.7). It is expected that mutation of *Hesx1*, *Hesx4*, *Hex*, *MMoxA* and *MMoxB* will result in disruption of the molecular mechanisms which mediate pluripotent stem cell differentiation and renewal during early embryogenesis. In addition, this approach may reveal additional functional roles of these genes during later stages of embryogenesis. Comparison between the null mutant phenotypes generated by mutation of these homeobox genes will also provide valuable insight into distinct developmental roles of these genes in the pre-gastrulation embryo, and will aid in the understanding of the developmental role of homeobox genes in general.

REFERENCES

Adamson, E. D. (1988) In *Experimental approaches to mammalian embryonic development,* Cambridge University Press, Cambridge pp. 321-364

Adeniyi-Jones, S. and Zasloff, M. (1985) *Nature* **317**, 81-84

Allen, J. D., Lints, T., Jenkins, N. A., Copeland, N. G., Strasser, A., Harvey, R. P. and Adams, J. M. (1991) *Genes & Dev.* **5**, 509-520

Altman, P. L. and Katz, D. D. (1979) *Inbred and genetically defined strains of laboratory animals.* Part I: *Mouse and rat.* Federation of American Societies for Experimental Biology, Bethesda, Maryland

Awgulwitsch, A. and Jacob, D. (1992) *Nature* **358**, 341-344

Baron, A., Featherstone, M. S., Hill, R. E., Hall, A., Galliot, B. and Duboule, D. (1987) *EMBO J.* **6**, 2977-2986

Bateson, W. (1894) *Materials for the Study of Variation*, MacMillan and Co., London

Beddington, R. S. P. (1982) *J. Embryol. exp. Morph.* **69**, 265-285

Beddington, R. S. P. (1988) In *Experimental approaches to mammalian development,* Cambridge University Press, Cambridge, pp 121-147

Beddington, R. S. P. and Robertson, E. J. (1989) *Development* **105**, 733-737

Bedford, F. K., Ashworth, A., Enver, T. and Wiedemann, L. M. (1993) *Nuc. Acids. Res.* **21**, 1245-1249

Ben-David, L., Aberdam, D., Sachs, L. and Blatt, C. (1991) *Virology* **182**, 382-387

Bennett, K. L., Hill, R. E., Pietras, M., Woodworth-Gutai, M., Kane-Haas, C., Houston, J. M., Heath, J. K. and Hastie, N. D. (1984) *Mol. Cell. Biol.* **4**, 1561-1571

Bettess, M. D. (1993) Honours Thesis, Department of Biochemistry, University of Adelaide

Bienz, M. and Tremml, G. (1988) *Nature* **333**, 576-578

Blochlinger, K., Bodmer, R., Jack, J., Jan, L. Y. and Jan, Y. N. (1988) *Nature* **333**, 629-635

Blum, M., Gaunt, S. J., Cho, K. W. Y., Steinbeisser, H., Blumberg, B., Bittner, D. and De Robertis, E. M. (1992) *Cell* **69**, 1097-1106

Bober, E., Baum, C., Braun, T. and Arnold, H-H. (1994) *Dev. Biol.* **162**, 288-303

Boncinelli, E., Simeone, A., Acampora, D. and Mavilio, F. (1991) *TIG* **7**, 329-334

Boncinelli, E., Somma, R., Acampora, D., Pannese, M., D'Esposito, M., Faiella, A. and Simeone, A. (1988) *Hum. Reprod.* **3**, 880-886

Bourbon, H-M., Lapeyre, B. and Amalric, F. (1988) *J. Mol. Biol.* **200**, 627-638

Bradley, I., Evans, M. Kaufman, M. H. and Robertson, E. J. (1984) *Nature* **309**, 255-256

Brennan, R. G. and Matthews, B. W. (1989) *J. Biol. Chem.* **264**, 1903-1906

Burglin, T. (1994) In *A guidebook to homeobox genes*, Oxford University Press, Oxford

Capecchi, M. R. (1989) *Science* **244**, 1288-1292

Carroll, S. B., Laymon, R. A., McCutcheon, M. A., Riley, P. D. and Scott, M. P. (1986) *Cell* **47**, 113-122

Chen, C. and Okayama, H. (1987) *Mol. Cell. Biol.* **7**, 2745-2752

Cho, K. W. Y., Blumberg, B., Steinbeisser, H. and De Robertis, E. M. (1991) *Cell* **67**, 1111-1120

Chomczynski, P. and Sacchi, N. (1987) *Analytical Biochem.* **162**, 156-159

Colberg-Poley, A. M., Voss, S. D., Chowdhury, K. and Gruss, P. (1985a) *Nature* **314**, 713-718

Colberg-Poley, A. M., Voss, S. D., Chowdhury, K., Stewart, C. L., Wagner, E. F. and Gruss, P. (1985b) *Cell* **43**, 39-45

Condie, B. G. and Capecchi, M. R. (1993) *Development* **119**, 579-595

Conover, J. C., Ip, N. Y., Poueymirou, W. T., Bates, B., Goldfarb, M. P., DeChiara, T. M. and Yancopoulos, G. D. (1993) *Development* **119**, 559-565

Davidson, D., Graham, E., Sime, C. and Hill, R. (1988) *Development* **104**, 305-316

Davis, C. A. and Joyner, A. L. (1988) *Genes & Dev.* **2**, 1736-1744

De Jong, R. and Meijlink, F. (1993) *Dev. Biol.* **157**, 133-146

Dietrich, W., Katz, H., Lincoln, S. E., Shin, H-S., Friedman, J., Dracopoli, N. C. and Lander, E. S. (1992) *Genetics* **131**, 423-447

Diwan, S. and Stevens, L. C. (1974) *Mouse News Lett.* **51**, 24-25

Doetschmann, T. C., Eistetter, M., Katz, M., Schmidt, W. and Kemler, R. J. (1985) *J. Embryol Exp. Morph.* **87**, 27-45

Dolle, P., Dierich, A., LeMeur, M., Schimmang, T., Schuhbaur, B., Chambon, P. and Duboule, D. (1993) *Cell* **75**, 431-441

Dolle, P., Izpisua-Belmonte, J-C., Brown, J., Tickle, C. and Duboule, D. (1991) *Genes Dev.* **5**, 1767-1776

Duboule, D. and Dolle, P. (1989) *EMBO J.* **8**, 1497-1505

Edwards, D. R., Parfett, C. L. J. and Denhardt, D. (1985) *Mol. Cell. Biol.* **5**, 3280-3288

Epstein, D. J., Vekemans, M. and Gros, P. (1991) *Cell* **67**, 767-774

Evans, M. J. and Kaufman, M. H. (1981) *Nature* **292**, 154-156

Fortini, M. E., Lai, Z. and Rubin, G. M. (1991) *Mech. Dev.* **34**, 113-122

Frischer, L. E., Hagen, F. S. and Garber, R. L. (1986) *Cell* **47**, 1017-1023

Frohman, M. A., Boyle, M. and Martin, G. R. (1990) *Development* **110**, 589-607

Frohman, M. A., Dush, M. K. and Martin, G. R. (1988) *Proc. Natl. Acad. Sci.* **85**, 8998-9002

Gaunt, S. J. (1991) *BioEssays* **13**, 505-513

Gaunt, S., Sharpe, P. T. and Duboule, D. (1988) *Development* **104** (Suppl.), 169-179

Giampaolo, A., Acampora, D., Zappavinga, V., Pannese, M., D'Esposito, M., Care, A., Faiella, A., Stornaiuolo, A. and Russo, G. (1989) *Differentiation* **40**, 191-197

Gilbert, W. (1985). *Science* **228**, 823-824

Glasser, S. W., Korfhagen, T. R., Bruno, M. D., Dey, C. and Whitsett, J. A. (1990) *J. Biol. Chem.* **265**, 21986-21991

Gould, A. P. and White, R. A. H. (1992) *Development* **116**, 1163-1174

Goulding, M. D., Chalepakis, G., Deutsch, U., Erselius, J. and Gruss, P. (1991) *EMBO J.* **10**, 1135-1147

Graham, A., Papalopulu, N. and Krumlauf, R. (1989) *Cell* **57**, 367-378

Green, M. C. (1972) *Mouse News Lett.* **47**, 36

Gruss, P. and Walther, C. (1992) *Cell* **69**, 719-722

Guthrie, S., Muchamore, I., Kuriowa, A., Marshall, H., Krumlauf, R. and Lumsden, A. (1992) *Nature* **356**, 157-159

Hafen, E., Levine, M. and Gehring, W. J. (1984) *Nature* **307**, 287-289

Hamada, H. and Kakunaga, T. (1982) *Nature* **298**, 396-398

Han, K., Levine, M. S. and Manley, J. L. (1989) *Cell* **56**, 573-583

Handyside, A. H., O.Neil, G. T., Jones, M. and Hooper, M. L. (1989) *Roux's Arch. Dev. Biol.* **198**, 8-55

Hastie, N. D. (1989) In *Genetic Variants and Strains of the Laboratory Mouse*, Oxford University Press, Oxford, pp 559-573

Hatano, M., Roberts, C. W. M., Minden, M., Crist, W. M. and Korsmeyer, S. J. (1991) *Science* **253**, 79-82

Hayashi, S. and Scott, M. P. (1990) *Cell* **63**, 883-894

Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., Ruvkun, G. and Horvitz, H. R. (1988) *Genes & Dev.* **2**, 1513-1516

Hill, R. E., Favor, J., Hogan, B. L. M., Ton, C. C. T., Saunders, G. F., Hanson, I. M., Prosser, J., Jordan, T., Hastie, N. D. and van Heyningen, V. (1991) *Nature* **354**, 522-525

Hill, R. E., Jones, P. F., Rees, A. R., Sime, C. M., Justice, M. J., Copeland, N. G., Jenkins, N. A., Graham, E. and Davidson, D. R. (1989) *Genes & Dev.* **3**, 26-37

Hogan, B., Constantini, F. and Lacy, E. (1986) *Manipulating the Mouse Embryo: a Laboratory Manual*, Cold Spring Harbor Laboratory, USA

Holland, P. W. H. (1991) *Gene* **98**, 253-257

Holland, P. W. H. (1992) *BioEssays* **14**, 267-273

Hooper, M., Hardy, K., Handyside, A., Hunter, S. and Monk, M. (1987) *Nature* **326**, 292-295

Hunt, P. and Krumlauf, R. (1992) *Annu. Rev. Cell Biol.* **8**, 227-256

Hunt, P., Gulisano, M., Cook, M., Sham, M-H., Faiella, A., Wilkinson, D., Boncinelli, E. and Krumlauf, R. (1991) *Nature* **353**, 861-864

Izpisua-Belmonte, J-C. and Duboule, D. (1992) *Dev. Biol.* **152**, 26-36

Izpisua-Belmonte, J-C., Tickle, C., Dolle, P., Wolpert, L. and Duboule, D. (1991) *Nature* **350**, 585-589

Jagla, K., Georgel, P., Bellard, F., Dretzen, G. and Bellard, M. (1993) *Gene* **127**, 165-171

Jaynes, J. B. and O'Farrell, P. H. (1988) *Nature* **336**, 744-749

Jeannotte, L., Lemieux, M., Charron, J., Poirier, F. and Robertson, E. J. (1993) *Genes & Dev.* **7**, 2085-2096

Jegalian, B. G. and De Robertis, E. M. (1992) *Cell* **71**, 901-910

Johnson, B. V. (1992) Honours Thesis, Department of Biochemistry, University of Adelaide

Johnson, M. H. and Maro, B. (1988) In *Experimental approaches to mammalian development*, Cambridge University Press, Cambridge, 35-65

Joyner, A. L. and Martin, G. R. (1987) *Genes & Dev.* **1**, 29-38

Kamb, A., Weir, M., Rudy, B., Varmus, H. and Kenyon, C. (1989) *Proc. Natl. Acad. Sci.* **86**, 4372-4376

Kamps, M. P., Murre, C., Sun, X. and Baltimore, D. (1990) *Cell* **60**, 547-555

Kappen, C., Schughart, K. and Ruddle, F. H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5459-5463

Kappen, C., Schughart, K. and Ruddle, F. H. (1993) *Genomics* **18**, 54-70

Kaptein, R. (1991) *Curr. Opin. Struct. Biol.* **1**, 63-70

Kaufman, M. H. (1992) *The Atlas of Mouse Development*, Academic Press Limited, London

Kelly, J. M., Gilbert, C. S., Stark, G. R. and Kerr, I. M. (1985) *Eur. J. Biochem.* **153**, 367-371

Kelly, J. M., Porter, A. C. G., Chernajovosky, Y., Gilbert, C. S., Stark, G. R. and Kerr, I. M. (1986) *EMBO J.* **7**, 1601-1606

Kennedy, M. A., Gonzalez-Sarmiento, R., Kees, U. R., Lampert, F., Dear, N., Boehm, T. and Rabbitts, T. H. (1991) *Proc. Natl. Acad. Sci.* **88**, 8900-8904

Kerppola, T. K. and Curran, T. (1991) *Curr. Opin. Struct. Biol.* **1**, 71-79

Kessel, M and Gruss, P. (1990) *Science* **249**, 374-379

Kessel, M., Balling, R. and Gruss, P. (1990) *Cell* **61**, 301-308

Kidder, G. M. (1992) *Dev. Genet.* **13**, 319-325

Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. and Pabo, C. O. (1990) *Cell* **63**, 579-590

Kondo, T., Takahashi, N. and Muramatsu, M. (1992) *Nucleic Acids Res.* **20**, 5729-5735

Kranz, D. M., Saito, H., Disteche, C. M., Swisshelm, Pravtcheva, D., Ruddle, F. H., Eisen, H. N. and Tonegawa, S. (1985) *Science* **227**, 941-945

Krasnow, M. A., Saffman, E. E., Kornfeld, K. and Hogness, D. S. (1989) *Cell* **57**, 1031-1043

Krieg, P. A. and Melton, D. A. (1987) *Meth. Enz.* **155**, 397-415.

Krumlauf, R. (1993a) *TIG* **9**, 106-112 .

Krumlauf, R. (1993b) *Current Opinion in Genetics and Development* **3,** 621-625

Kuehn, M. R., Bradley, A., Robertson, E. J. and Evans, M. R. (1987) *Nature* **326**, 255-256

Latham, K. E., Garrels, J. I., Chang, C. and Solter, D. (1991) *Development* **112**, 921-932

Le Mouellic, H., Lallemand, Y. and Brulet, P. (1992) *Cell* **69**, 251-264

Lewis, E. (1978) *Nature* **276**, 565-570

Liang, P. and Pardee, A. B. (1992) *Science* **257**, 967-970

Lints, T. J., Parsons, L. M., Hartley, L., Lyons, I. and Harvey, R. P. (1993) *Development* **119**, 419-431

Lobe, C. G. and Gruss, P. (1989) *New Biologist* **1**, 9-18

Lu, M., Gong, Z., Shen, W. and Ho, A. D. (1991) *EMBO J.* **10**, 2905-2910

Lufkin, T., Mark, M., Hart, C.P., Dolle, P., LeMeur, M. and Chambon, P. (1992) *Nature* **359**, 835-841

Mahon, K. A., Westphal, H. and Gruss, P. (1988) *Development* **104** (Suppl.), 187-195

Mark, M., Lufkin, T., Vonesch, J-L., Ruberte, E., Olivo, J-C., Dolle, P., Gorry, P., Lumsden, A. and Chambon, P. (1993) *Development* **119**, 319-338

Maro, Gueth-Hallonet, Aghion and Antony (1991) *Development* (Suppl.) 1, 17-25

Martin, G. R. (1975) *Cell* **6**, 229-243

Martin, G. R. (1980) *Science* **209**, 768-776

Martin, G. R. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7634-7638

Martin, G. R. and Evans, M. J. (1975) in *Teratomas and Differentiation*, 169-187, Academic Press, New York, 169-187

Matsui, T., Hirai, M., Hirano, M. and Kurosawa, Y. (1994) *FEBS Lett.* **336**, 107-110

Matsui, T., Hirai, M., Wakita, M., Hirano, M. and Kurosawa, Y. (1993) *FEBS Lett.* **322**, 181-185

Matsui, Y., Zsebo, K. and Hogan, B. L. M. (1992) *Cell* **70**, 841-848

Mavilio, F., Simeone, A., Boncinelli, E. and Andrews, P. W. (1988) *Differentiation* **37**, 73-79

McGinnis, W. and Krumlauf, R. (1992) *Cell* **68**, 283-302

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. and Gehring, W. J. (1984b) *Cell* **37**, 403-408

McGinnis, W., Hart, C. P., Gehring, W. J. and Ruddle, F. H. (1984c) *Cell* **38**, 675-680

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. and Gehring, W. J. (1984a) *Nature* **308**, 428-433

Meijer, D., Graus, A., Kraay, R., Langeveld, A., Mulder, M. P. and Grosveld, G. (1990) *Nuc. Acids Res.* **18**, 7357-7365

Millen, K. J., Wurst, W., Herrup, K. and Joyner, A. L. (1994) *Development* **120**, 695-706

Monaghan, A. P., Davidson, D. R., Sime, C., Graham, E., Baldock, R., Bhattacharya, S. S. and Hill, R. E. (1991) *Development* **112**, 1053-1061

Monk, M. and McLaren, A. (1981) *J. Embryol. Exp. Morphol.* **63**, 75-84

Moore, K. J., D'Amore-Bruno, M. A., Korfhagen, T. R., Glasser, S. W., Whitsett, J. A., Jenkins, N. A. and Copeland, N. G. (1992) *Genomics* **12**, 388-393

Morgan, B. A. and Tabin, C. J. (1993) *Current Opinion in Genetics and Development* **3**, 668-674

Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459-472

Mountford, P., Zevnik, B., Duwel, A., Nichols, J., Li, M., Dani, C., Robertson, M., Chambers, I. and Smith, A. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 4303-4307

Murphy , P. and Hill, R. E. (1991) *Development* **111**, 61-74

Murtha, M. T., Leckman, J. F. and Ruddle, F. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10711-10715

Nadeau, J. H. and Cox, R. (1992) *Mamm. Genome* **3**, S206-S219

Nesbitt, M. N. and Francke, U. (1973) *Chromosoma* **41**, 145-158

Neufeld, E. J., Skalnik, D. G., Lievens, M-J. Orkin, S. H. (1992) *Nature Genet.* **1**, 50-55

Okamoto, K., Okazowa, H., Okuda, A., Sakai, M., Muramatsu, M. and Hamada, H. (1990) *Cell* **60**, 461-472

Opstelten, D-J. E., Vogels, R., Robert, B., Kalkhoven, E., Zwartkruis, F., de Laaf, L., Destree, O. H., Deschamps, J., Lawson, K. A. and Meijlink, F. (1991) *Mech. of Dev.* **34**, 29-42

Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J. and Wuthrich, K. (1990) *EMBO J.* **9**, 3085-3092

Papalopulu, N., Hunt, P., Wilkinson, D., Graham, A. and Krumlauf, R. (1990) *Adv. Neural Regeneration Res.* 291-307

Papalopulu, N., Lovell-Badge, R. and Krumlauf, R. (1991) *Nucleic Acids Res.* **19**, 5497-5506

Pearson, W. R. and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444-2448

Pease, S., Braghetta, P., Gearing, D., Grail, D. and Williams, R. L. (1990) *Dev. Biol.* **141**, 344-352

Pederson, R. A. (1988) In *Experimental approaches to mammalian development*, Cambridge University Press, Cambridge, 3-33

Perkins, A. C. and Cory, S. (1993) *EMBO J.* **12**, 3835-3846

Perkins, A. C., Kongsuwan, K., Visvader, J., Adams, J. M. and Cory, S. (1990) *Proc. Natl. Acad. Sci.* **87**, 8398-8402

Phillips, M., Djian, P. and Green, H. (1986) *J. Biol. Chem.* **261**, 10821-10827

Pick, L., Schier, A., Affolter, M., Schmidt-Glenewinkel, T. and Gehring, W. J. (1990) *Genes & Dev.* **4**, 1224-1239

Piquet-Pellorce, C., Grey, L., Mereau, A. and Heath, J. K. (1994) *Exp. Cell Res.* **213**, 340-347

Popperl, H. and Featherstone, M. S. (1993) *Mol. Cell. Biol.* **13**, 257-265

Price, M., Lemaistre, M., Pischetola, M., Di Lauro, R. and Duboule, D. (1991) *Nature* **351**, 748-751

Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J. and Wuthrich, K. (1989) *Cell* **59**, 573-580

Rabbitts, T. H. and Boehm, T. (1990) *Nature* **346**, 418

Ramirez-Solis, R., Zheng, H., Whiting, J., Krumlauf, R. and Bradley, A. (1993) *Cell* **73**, 279-294

Rathjen, P. D., Nichols, J., Toth, S., Edwards, D. R., Heath, J. K. and Smith, A. G. (1990) *Genes & Dev.* **4**, 2308-2318

Roberts, C. W. M., Shutter, J. R. and Korsmeyer, S. J. (1994) *Nature* **368**, 747-749

Robertson, E. J. (1987) in *Teratocarcinomas and Embryonic Stem Cells, A Practical Approach*, IRL Press, Oxford, pp. 71-112

Robertson, E. J. and Bradley, A. (1988) In *Experimental approaches to mammalian embryonic development,* Cambridge University Press, Cambridge pp. 475-508

Robertson, M., Chambers, I., Rathjen, P., Nichols, J. and Smith, A. (1993) *Dev. Genet.* **14**, 165-173

Rosner, M. H., Vigano, M. A., Ozato, K., Timmons, P. M., Poirier, F., Rigby, P. W. J. and Staudt, L. M. (1990) *Nature* **345**, 686-692

Rosner, M. H., Vigano, M. A., Rigby, P. W. J., Arnheiter, H. and Staudt, L. M. (1991) *Science* **253**, 144-145

Rossant, J. (1993) *Current Biol.* **3**, 47-49

Saksela, K. and Baltimore, D. (1993) *Mol. Cell. Biol.* **13**, 3698-3705

Samson, M-L., Jackson-Grusby, L. and Brent, R. (1989) *Cell* **57**, 1045-1052

Sasaki, A. W., Doskow, J., MacLeod, C. L., Rogers, M. B., Gudas, L. J. and Wilkinson, M. F. (1991) *Mech. of Dev.* **34**, 155-164

Schier, A. F. and Gehring, W. J. (1992) *Nature* **356**, 804-807

Schmid, C. W. and Jelinek, W. R. (1982) *Science* **216**, 1065-1070

Scholer, H. R. (1991) *TIG* **7**, 323-329

Scholer, H. R., Dressler, G. R., Balling, R., Rohdewohld, H. and Gruss, P. (1990b) *EMBO J.* **9**, 2185-2195

Scholer, H. R., Hatzopolous, A. K., Balling, R., Suzuki, N. and Gruss, P. (1989) *EMBO J.* **8**, 2543-2550

Scholer, H. R., Ruppert, S., Suzuki, N., Chowdhury, K. and Gruss, P. (1990a) *Nature* **344**, 435-439

Schughart, K., Kappen, C. and Ruddle, F. H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7067-7071

Schultz, R. M. (1988) In *Experimental approaches to mammalian embryonic development,* Cambridge University Press, Cambridge pp.195-237

Schwartzberg, P., Goff, S. P. and Robertson, E. J. (1989) *Science* **246**, 799-802

Scott, M. P. (1992) *Cell* **71**, 551-553

Scott, M. P. and Weiner, A. J. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 4115-4119

Scott, M. P., Tamkun, J. W. and Hartzell, G. W., III (1989) *Biochim. Biophys. Acta* **989**, 25-48

Shimeld, S. M. and Sharpe, P. T. (1992) *Biochim. Biophys. Acta* **1136**, 253-258

Simeone, A, Acampora, D., Nigro, V., Faiella, A., D'Esposito, M., Stornaiuolo, A., Mavillo, F. and Boncinelli, E. (1991) *Mech. Dev.* **33**, 169-248

Simeone, A., Acampora, D., Arcioni, L., Andrews, P., Boncinelli, E. and Mavillo, F. (1990) *Nature* **346**, 763-766

Simeone, A., Acampora, D., Massimo, G., Stornaiuolo, A. and Boncinelli, E. (1992) *Nature* **358**, 687-690

Singer, M. F. (1982) *Cell* **28**, 433-434

Singh, G., Kaur, S., Stock, J. L., Jenkins, N. A., Gilbert, D. J., Copeland, N. G. and Potter, S. S. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10706-10710

Smith, A. G. (1991) *J. Tiss. Cult. Meth.* **13**, 89-94

Smith, A. G. (1992) *Seminars in cell biology* **3**, 385-399

Smith, A. G. and Hooper, M. L. (1987) *Dev. Biol.* **121**, 1-9

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D. (1988) *Nature* **336**, 688-690

Snow, M. H. L. (1981) *J. Embryol. Exp. Morphol.*. (Suppl) **65**, 269-287

Snow, M. H. L. and Monk, M. (1983) In *Current problems in germ cell differentiation*, Cambridge University Press, Cambridge, pp 115-136

Snow, M. H. L. and Tam, P. P. L. (1979) *Nature* **279**, 554-557

Solter, D. and Knowles, B. B. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5565-5569

Spemann, H. and Mangold, H. (1924) *Roux' Arch. Entw. Mech.* **100**, 599-638

Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hildebrand, C. E. and Moyzis, R. K. (1991) *Genomics* **10**, 807-815

Stewart, T. A. and Mintz, B. (1981) *Proc. Natl. Acad. Sci.* **78**, 6314-6318

Suzuki, N., Rohdewohld, H., Neuman, T., Gruss, P. and Scholer, H. R. (1990) *EMBO J.* **9**, 3723-3732

Taga, T., Narazaki, M., Yasukawa, K., Saito, T., Miki, D., Hamaguchi, M., Davis, S., Shoyab, M., Yancopolous, G. D. and Kishimoto, T. (1992) *Proc. Natl. Acad. Sci.* **89**, 10998-11001

Tam, P. P. L. and Beddington, R. S. P. (1992) In *Postimplantation development in the mouse*, John Wiley & Sons, London, pp. 27-49

Tam, P. P. L. and Snow, M. H. L. (1981) *J. Embryol. Exp. Morphol.*. **64**, 133-147

Tarkowski, A. K. (1959) *Nature* **184**, 1286-1287

Thomas, P. Q. and Rathjen, P. D. (1992) *Nucleic Acids Res.* **20**, 5840

Thomas, P. Q., Johnson, B. V., Rathjen, J. and Rathjen, P. D. (1994) *J. Biol. Chem.*, In Press

Tonissen, K. F., Drysdale, T. A., Lints, T. J., Harvey, R. P. and Krieg, P. A. (1994) *Dev. Biol.* **162**, 325-328

Treisman, J., Harris, E. and Desplan, C. (1991) *Genes & Dev.* **5**, 594-604

Valarche, I., Tissier-Seta, J-P., Hirsch, M-R., Martinez, S., Goridis, C. and Brunet, J-F. (1993) *Development* **119**, 881-896

Walter, H., Brooks, D. and Fisher, D. (1985) *Partitioning in Aqueous Two-phase Systems*, Academic Press, London

Walter, P. and Blobel, G. (1982) *Nature* **299**, 691-698

Wang, R., Clark, R. and Bautch, V. L. (1992) *Development* **114**, 303-316

Way, J. C. and Chalfie, M. (1988) *Cell* **54**, 5-16

Weber, J. L. (1990) *Genomics* **7**, 524-530

Whyatt, L. M., Duwel, A., Smith, A. G. and Rathjen, P. D. (1993) *Mol. Cell. Biol.* **13**, 7971-7976

Wilkinson, D., Bhatt, S., Cook, M., Boncinelli, E. and Krumlauf, R. (1989) *Nature* **341**, 405-409

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A. and Gough, N. M. (1988) *Nature* **336**, 684-687

Winslow, G. M., Hayashi, S., Krasnow, M., Hogness, D. S. and Scott, M. P. (1989) *Cell* **57**, 1017-1030

Wright, C. V. E. (1993) *Current Biology* **3**, 618-621

Yeom, Y. I., Ha, H-S., Balling, R., Scholer, H. and Artzt, K. (1991) *Mech. of Dev.* **35**, 171-179

Young, P. R., Scott, R. W., Hamer, D. H. and Tilghman, S. M. (1982) *Nuc. Acids. Res.* **10**, 3099-3116

Zaraisky, A. G., Lukyanov, S. A., Vasiliev, O. L., Smirnov, Y. V., Belyavsky, A. V. and Kazanskaya, O. V. (1992) *Dev. Biol.* **152**, 373-382

Zimmermann, J. W. and Schultz, R. M. (1994) *Proc. Natl. Acad. Sci.* **91**, 5456-5460