ANALYSIS OF MURINE HOMEBOX GENES

ANTHONY GRAHAM

Thesis for the Degree of Doctor of Philosophy

Division of Eukaryotic Molecular Genetics
National Institute for Medical Research
The Ridgeway, Mill Hill
London NW7 1AA

University College London
Gower Street
London WC1E 6BT

November, 1989
This thesis is dedicated to my mother and father.
The work in this thesis is original except where stated. The sequences used for comparisons in Chapter 3 were not determined by the author but were either from publications from other laboratories or from other members of the laboratory.

Anthony Graham:
The work in this thesis is consistent with, and strengthens, suggestions that murine homeobox genes play a role in the establishment of regional specification. It is shown that there is a correlation between the physical order of members of the Hox 2 cluster and their order of expression along the rostrocaudal axis, such that each more 3' gene is expressed more rostrally. Such a correlation has been previously shown for the *Drosophila* homeotic genes of the ANT-C and BX-C, where it was clear that these genes were involved in the establishment of regional identity along the rostrocaudal axis. It has also been demonstrated that the relationship between the murine and *Drosophila* clusters extends further than aspects of their expression patterns. Sequence comparisons suggest that the murine Hox clusters and the *Drosophila* homeobox clusters arose from a common ancestral cluster that existed prior to the divergence between the protostomes and the deuterostomes. Therefore the corresponding utilisation of ordered region specific expression of members of homeobox clusters in flies and mice together with their conservation would suggest that these genes are involved in the establishment of rostrocaudal regional specification in these organisms, and by inference in all other higher metazoans. It is also demonstrated that the Hox 2 genes show spatially and temporally dynamic patterns of expression in the transverse plane of the developing central nervous system. Since all of the Hox 2 genes exhibit similar expression patterns it is felt they are reflecting events during ontogeny of the spinal cord. The observed patterns correlate with the timing and location of the birth of the major classes of neurons. It is suggested that the Hox 2 genes are conferring rostrocaudal positional information on each class of newly born neurons. These genes also exhibit a striking dorsal restriction in their expression within the developing spinal cord which does not appear to correlate with morphology and eventually breaks down.
ACKNOWLEDGEMENTS

Firstly, I would like to thank my external supervisor Dr. Robb Krumlauf for all of his excellent help, encouragement and guidance during the course of this work. I also thank Dr. David Latchman (University College, London) for efficiently fulfilling his role as my University supervisor. This research was carried out at the National Institute for Medical Research and was supported by the MRC. I wish to acknowledge the assistance provided by the NIMR photography section.

I also thank many of the past and present members of NIMR, particularly of the laboratories of Molecular Embryology and Embryogenesis, for all of their help and friendship. Particular thanks to a few gurus - Jonathan Cooke, Peter Holland, Malcolm Maden and Dennis Summerbell who have been an invaluable source of information and discussion. Jeremy Green and Peter Vize have been wonderful proof readers, spotting many a ridiculous mistake, and who have also come up with more than a few smart ideas. Extra thanks go to Nancy Papalopulu, my long suffering lab mate and friend, who has helped immensely throughout the period of this work both technically and intellectually and without whom I don’t know what I would have done. I also wish to thank Pal Njolstad, the zebrafish man, who has generously supplied much unpublished information.

Cheers Big Man, by the way - Sheila, Kathleen, Gerard, Philip, Brian, Carol and Dan and not to forget Sandy for many an interesting biological discussion. Finally thanks to mum and dad for everything for the past many years.
CONTENTS

Title Page

Dedication

Abstract

Acknowledgements

Contents

List of Figures

Abbreviations

References

Chapter One : Introduction

1.1 Regional specification.

1.2 A summary of mouse embryogenesis.

1.2.1 Pre-implantation development.

1.2.2 Gastrulation and the generation of the body plan.

1.3 Strategies towards a molecular understanding of murine development.

1.4 Regional specification in Drosophila embryogenesis.

1.4.1 The genetics of anteroposterior specification.

1.4.2 The genetics of dorsoventral specification.

1.4.3 Neurogenesis.

1.5 The homeobox.

Chapter Two : Material and Methods
2.1 Buffers and stock solutions. 19
2.2 Bacterial media and antibiotics. 22
2.3 Animals. 22
2.4 Cloning and DNA methods. 22
2.4.1 Bacterial strain. 22
2.4.2 The production of cells competent for transformation. 23
2.4.3 Bacterial transformation. 23
2.4.4 Standard DNA methods. 23
2.4.5 Restriction digests. 24
2.4.6 Agarose gel electrophoresis. 24
2.4.7 Isolation of DNA fragments from gels. 24
2.4.8 Sub-cloning and ligations. 24
2.4.9 Isolation of plasmid DNA- mini prep. 25
2.4.10 Isolation of plasmid DNA- maxi prep. 25
2.4.11 Southern blot. 26
2.4.12 Colony screen. 26
2.5 RNA methods. 27
2.5.1 In vitro transcription. 27
2.5.2 Agarose formaldehyde gels. 28
2.5.3 Northern blot. 28
2.5.4 In situ hybridisation. 28
2.6 Sequence comparisons. 30
2.6.1 Sequences used for comparisons. 30
2.7 The Hox 3.1 probe 31

Chapter Three: The murine and Drosophila clusters are evolutionary related. 32
3.1 Introduction. 32
3.2 The mouse Hox clusters are related, possibly by duplication. 35
3.3 Similarities between mouse Hox complexes and those of other vertebrates.

3.4 The mouse and Drosophila homeobox gene complexes have common features of organisation.

3.5 A common ancestor for the mouse and Drosophila homeobox gene complexes.

Chapter 4: Expression of Hox 2 genes along the rostrocaudal axis of the mouse embryo.

4.1 General features of murine homeobox gene expression.

4.2 Relationship between the clustering of homeoboxes and their expression.

4.3 The Hox 2 probes used in in situ hybridisation.

4.4 Ordered expression of Hox 2 genes in the central nervous system (CNS).

4.5 Correlation between gene order and expression is also evident in other ectodermal derivatives.

4.6 Ordered differential expression of Hox 2 genes in mesodermal derivatives.

4.7 Discussion: Position of genes within a cluster reflect domains of expression along the rostrocaudal axis for both mouse and Drosophila.

4.7.1 Ordered expression in ectoderm.

4.7.2 Hox 2 gene expression in the hindbrain and rhombomeres.

4.7.3 Hox 2 gene expression in the mesoderm.

4.7.4 The anterior boundaries in mesoderm and ectoderm are not coincident.

4.7.5 Correlation between gene order and expression is a general aspect of homeobox clusters.
Chapter Five: Dynamic Hox 2 gene expression during development of the central nervous system.

5.1 Introduction

5.2 Hox 2 genes are dorsally restricted at 12.5 days

5.3 Hox 2 genes are expressed uniformly across the nerve cord at 10.5 days.

5.4 Strong lateral expression of Hox 2 genes at 11.5 days.

5.5 Expression of Hox 2 genes reappears ventrally at 14.5 days.

5.6 Hox 3.1 is expressed in stripes.

5.7 Discussion

5.7.1 Hox 2 gene expression correlates with the time and place of the birth of the major classes of spinal neurons.

5.7.2 The Hox 2 genes and Hox 3.1 must be responding to different signals.

5.7.3 Re-expression ventrally and the sharp dorsal demarcation in the expression patterns.

Chapter Six: General discussion

6.1 Positions of genes in a cluster reflect domains of expression in the rostrocaudal axis for both mouse and Drosophila.

6.2 Evolution of homeobox gene clusters and the evolution of the body plan.

6.3 Relationships between the expression patterns of genes from related murine Hox clusters.

6.4 Hox 2 gene expression and the vertebrate head.
6.5 Hox genes are involved in the interpretation of regional cues.

6.6 Differences between Drosophila and mouse development that could preclude the use of identical genetic mechanisms.

6.7 Molecules that could organise Hox gene expression in vertebrates.

6.7.1 Growth factors and regional specification in Xenopus embryogenesis.

6.7.2 Retinoic acid and rostrocaudal regional specification.

6.7.3 Retinoic acid and homeobox gene expression.

6.8 Other roles for homeobox genes.

6.9 Homeobox target genes.

6.10 Future experimental approaches to homeobox gene function.

6.10.1 Tissue grafts and Hox gene expression.

6.10.2 Generation of gain of function mutants.

6.10.3 Generation of loss of function mutants.
LIST OF FIGURES

Chapter 3

3.1 - The organisation of the Hox 2 cluster of 9 murine homeobox genes and a diagrammatic representation of the Hox 2.6 gene structure.

3.2 - Gap alignment to identify regions of identity between different protein sequences.

3.3 - Amino acid sequence comparisons of the mouse and Drosophila homeodomains showing the existence of related subgroups.

3.4 - A schematic alignment of the Hox 1 and Hox 2 clusters.

3.5 - Comparison between three related murine homeodomain proteins.

3.6 - Generalised structure showing characteristic features and conserved regions of a typical vertebrate homeodomain protein.

3.7 - Diagrammatic representation of the relationship between Hox 2.6 and the Drosophila Dfd.

3.8 - Representation of the relationship between Drosophila and murine homeobox gene clusters.

3.9 - A possible scheme for the evolution of the Drosophila and murine clusters from a common ancestral cluster.

Chapter 4

4.1 - In situ hybridisation of the Hox 2.2 gene in the 12.5 day mouse embryo.

4.2 - Schematic representation of the fate map of a Drosophila blastoderm stage embryo.

4.3 - Comparisons of the patterns of expression for Hox 2.5, Hox 2.4 and Hox 2.2 in the 12.5 day mouse embryo.
4.4 - Anterior boundaries of expression in the central nervous system for six members of the Hox 2 cluster as revealed by \textit{in situ} hybridisation.

4.5 - Patterns of expression of the Hox 2 genes that show the most anterior limits of expression in the central nervous system.

4.6 - Diagrammatic representation of the anterior boundaries of expression in the central nervous system for the Hox 2 cluster, and their correlation with gene position in the cluster.

4.7 - Differential expression of members of Hox 2 in the vagal/glossopharyngeal nerve complex.

4.8 - Expression of Hox 2 genes in some neural crest derivatives.

4.9 - Anterior boundaries of expression of three members of Hox 2 in the prevertebrae.

4.10 - Differential expression of members of Hox 2 in the lung of the 12.5 day mouse embryo.

4.11 - Expression patterns of members of the Hox 2 cluster in the stomach and gut of a 14.5 day mouse embryo.

4.12 - Mesonephric tubules of the 10.5 day mouse embryo displaying expression of the Hox 2.4, Hox 2.2 and Hox 2.6 genes.

4.13 - Expression of the Hox 2 genes in the metanephric kidney, developing gonad and adrenal gland of a 14.5 day embryo.

Chapter 5

5.1 - Expression patterns of four members of Hox 2 at five different points along the rostrocaudal axis of a 12.5 day mouse embryo.

5.2 - The patterns of expression for seven members of the Hox 2 cluster
in the 10.5 day nerve cord.

5.3 - Changes in expression pattern of Hox 2.2 in the spinal cord between 10.5 and 11.5 days of gestation.

5.4 - Transverse sections of the 11.5 day, mouse nerve cord probed with Hox 2 genes.

5.5 - Expression of three Hox 2 genes at three different points along the rostrocaudal axis of the 14.5 day mouse embryo.

5.6 - A comparison between the expression patterns of the Hox 2.5 and the Hox 3.1 genes in the caudal 12.5 day mouse spinal cord.
ABBREVIATIONS

ANT-C  Antennapedia Complex
ATP  Adenosine triphosphate
BSA  Bovine Serum Albumin
BX-C  Bithorax Complex
cDNA  Complementary Deoxyribonucleic acid
CNS  Central Nervous System
DEPC  Diethyl Pyrocarbonate
DNA  Deoxyribonucleic acid
DTT  Dithiothreitol
EDTA  Ethylenediaminetetraacetic acid
ES  Embryonal Stem
EtBr  Ethidium Bromide
FGF  Fibroblast Growth Factor
ICM  Inner Cell Mass
b-Me  b-Mercaptoethanol
mRNA  Messenger Ribonucleic Acid
MOPS  Morpholinopropanesulfonic Acid
NaPB  Sodium Phosphate Buffer
O.D.  Optical Density
Poly A  Polyadenylic Acid
Poly C  Polycytidylic Acid
RNA  Ribonucleic Acid
SDS  Sodium Dodecyl Sulphate
SSC  Standard Sodium Citrate
TAE  Tris Acetate EDTA
TBE  Tris Borate EDTA
TE  Tris-EDTA
TESPA  3-Aminopropyltriethoxysilane
TGF  Transforming Growth Factor
TRIS  Tris(hydroxymethyl)methylamine
t-RNA  Transfer Ribonucleic acid
UTP  Uridine Triphosphate
XTC-MIF  XTC-Mesoderm inducing factor
1.1 Regional Specification

A core problem in development is that of regional specification: how the correct structures are formed in the correct place at the correct time. The process of cellular differentiation, which determines the synthesis of macromolecules that characterise a particular cell type, and that of cell movement, which is concerned with a change in form, are thought to be secondary (Wolpert, 1971). The role of regional specification would be in ensuring that the appropriate cellular differentiation occurs in the right place. For example, the cellular differentiation of the cartilage is most probably the same in the forelimb and hindlimb but its spatial organisation is different (Wolpert, 1971). This is also observed with the process of cell movement where again the importance of regional specification would be in determining what cells will move or generate the forces necessary for a given change in shape (Wolpert, 1971).

The body plan/parts of an animal are not specified all at once but are thought to result from a hierarchy of developmental decisions (Slack, 1983). It has been suggested that, during embryogenesis, at any given point there are a small number of decisions available to a cell, each of which is irreversible, under normal conditions. For example, for cells to be competent to form neural tissue they would have to decide to be ectoderm, rather than mesoderm or endoderm, and then to be neural rather than epidermal. It is felt that the outcome of any decision will not generally be a terminally differentiated state but rather a new state of determination which, if the decisions are irreversible, would result in a reduction in potency as compared with the previous state (Slack, 1983).

One could break regional specification into two independent processes: an instructive event during which the positional information is imparted, and, an interpretation by the competent tissue (Wolpert, 1971). Slack (1983) has suggested that there are two general processes which could account for the instructive signal: cytoplasmic localisation, and induction. Cytoplasmic localisation would be the process where, as the embryo divides, each of the daughter cells receive different cytoplasms and/or different membranes that cause them to undertake different paths of development. In contrast induction would involve, a competent tissue becoming differentially determined in response to a chemical signal or interaction with cells or matrix from other regions of the embryo. The interpretation would be the process through which the positional information is made apparent and would probably directly or

An important implication of regional specification is that cells of the same histological type could be in different states of determination i.e. they would be non-equivalent (Lewis and Wolpert, 1976). Again using the vertebrate limb as an example these authors have suggested that cells in the fore limb and hind limb can both differentiate as cartilage, however despite the similar cartilage phenotype the cells would be non-equivalent in terms of their positional values. It is suggested that this difference would eventually manifest itself in growth or surface properties such as between the wrist, and the radius and ulna (Lewis and Wolpert, 1976). The theory of non-equivalence also implies that cells carry internal records of past external influences and that these records may govern the response of cells to present external influences. In support of this Wolpert and Stein (1984) cite an experiment in which a piece of prospective thigh region was transplanted to the tip of the wing bud. This transplanted material developed into a toe and even induced the overlying ectoderm to form a covering of scales and claws. This experiment was interpreted as demonstrating that the transplanted material carried a leg label and remembered its history when it was transplanted to the wing bud. The nature of these records is thought of in terms of a gene control network. This network would confer identity on the cells through different combinations of genes which may be on or off (Lewis and Wolpert, 1976).

In terms of regional specification one could think of the differential activity of such a network as being part of the interpretation machinery of the cells which responded to the positional information. The record of non-equivalence that each cell possesses could be in the combinations of gene activity/inactivity which remain after the initial positional information has been imparted.

It is interesting to consider the evolutionary aspects of regional specification, such as the phylogenetic origin of the apparatus of regional specification and its degree of conservation. It is felt that there will be some degree of generality in the biochemical machinery underlying regional specification which may be in the positional signal(s) and/or in the apparatus of interpretation (Slack, 1983). A number of lines of evidence support this idea. Firstly, a mouse polarising region grafted to the anterior portion of a chick limb bud results in the specification of additional digits. It seems possible that the signal for organising the anteroposterior axis of the limb may be conserved in all amniotes (Fallon and Crosby, 1977). The neural tube which arises from dorsal ectoderm in all vertebrates is another case in that
there would seem to be evidence from both chick and amphibians that it is formed as a result of induction from the underlying mesoderm (Spemann, 1938, Grabowski, 1957).

These examples have dealt with vertebrates which represent only about 5% of the known species of animal and as such are a reasonably small sampling. It is interesting to speculate whether general mechanisms of regional specification would be conserved across much wider evolutionary distances. The similarities in early vertebrate development argue that there are highly conserved general mechanisms. However is there any common ground between insects and vertebrates? The comparisons between insects and vertebrates is intriguing since these animals represent the tips of the two major higher metazoan lineages, the protostomes and the deuterostomes respectively.

These two lineages derive their names from the origin of their definitive mouths. The protostomes derive their definitive mouth from the blastopore while the deuterostomes form their blastopore posteriorly and the mouth forms as a new structure (Balinsky, 1981). The protostomes tend to exhibit spiral cleavage which involves unequal cleavage and results in the separation of small upper micromeres from larger lower micromeres. This process of spiral cleavage is used to generate a fixed or predetermined cell fate (Balinsky, 1981). The deuterostomes exhibit radial rather than spiral cleavage. It is also generally stated that the protostomes and the deuterostomes form their coeloms in different manners (Clark, 1964). The former tending to form its coelom through cavitation of a solid mass of mesoderm. Such a coelom is termed a schizocoelom. The deuterostomes tend to from the mesoderm and the coelom together as pouches from the original gut cavity of the gastrula. The wall becomes the mesoderm and the separated cavity persists as the coelom. This type of coelom is called an enterocoelom.

Since the mouse and Drosophila represent two extremes in developmental strategies it would be interesting to compare them. Common aspects in their developmental processes would predict that any such process would have existed prior to the split between the protostomes and deuterostomes. By extrapolation they could be common to all higher metazoans, and provide important clues to general developmental strategies for regional specification.

1.2 A Summary of Mouse Embryogenesis

The work in this thesis has focused upon trying to identify genes that are important in vertebrate and particularly mammalian embryogenesis and for that purpose the mouse has been employed. The mouse is the best characterised vertebrate with regards to genetics, cell and molecular biology and therefore offers advantages that one does not find in other vertebrate systems. Through
reverse genetic manipulation of the mouse genome, using traditional transgenic technology or via embryonal stem (ES) cells, the capability now exists to alter the developmental patterns of gene expression by overexpressing or removing gene activity in the mouse embryo. The phenotypes that may result from these sort of experiments could provide approaches to critical processes and decisions in early embryogenesis. Thus the mouse offers many advantages to workers who are interested in the genetics of vertebrate and particularly mammalian development.

1.2.1 Pre-implantation development

The early stages of mouse development are not concerned with the specification of the body plan but with the segregation of extra-embryonic tissue from embryonic tissue. The extra-embryonic tissues are necessary for the support and nutrition of the embryo (Rugh, 1968, Theiler, 1972). The segregation of the cells which will produce the embryo depends on events during cleavage. There is evidence to suggest that at the 8-16 and 16-32 cell stage, cells differentiate according to their position such that the outer cells will be directed to form the trophectoderm, which is an extra-embryonic tissue, while the inner cells will contribute to the embryo. (Tarkowski and Wroblewska, 1967, Hillman et al., 1972). Hillman and co-workers demonstrated that while a four-cell blastomere could form both embryonic and extra-embryonic tissue if one placed a labelled four-cell blastomere on the outside of another four cell embryo then this cell tended to remain outside and to contribute to the extra-embryonic tissues.

At the eight cell stage the blastomeres of the mouse embryo undergo a dramatic shape change which is termed compaction wherein the spherical appearance the blastomeres becomes flattened (Lewis and Wright, 1935). The physical isolation of the cells, either inside or outside, is associated with compaction of the embryo at the eight- sixteen-cell stage (Ducibella and Anderson, 1975). This involves polarisation of the blastomeres of the embryo and the formation of tight junctions between cells at the periphery of the embryo (Ducibella and Anderson, 1975). The polarisation of the cells is demonstrable in many ways; localisation of microvilli to the apical surface, increased adhesiveness at the basolateral surface and nuclear displacement (Reeve, 1981a, Reeve and Ziomek, 1981b). The polarisation of the blastomeres precedes compaction and will occur if compaction is prevented (Ziomek and Johnson, 1980, Kimber and Surani, 1981). Elements of the polarity in the eight cell blastomere seem to be conserved through the next cell division (Johnson and Ziomek 1981). The asymmetry would be carried through cell division by orienting the plane of cleavage approximately perpendicular to the axis of
polarity. This can produce two 16-cell stage blastomeres one of which is polarised and has an apical surface with microvilli and while the other has no polarity and is adhesive at all points (Johnson et al., 1986).

The embryo is called a morula from compaction to the 32-cell stage when a fluid filled blastocoel begins to form in the interior (Rugh, 1968, Theiler, 1972). This activity expands the embryo into a blastocyst which contains an outer layer of cells, the trophectoderm, and a clump of cells attached to one point of its interior, the inner cell mass (ICM). The outside polar cells of the sixteen cell stage are those that contribute to the trophectoderm while those that are internal and apolar produce the progenitors of the inner cell mass (Tarkowski and Wroblewska, 1967). The cells of the ICM are still totipotent at this stage and give rise to both embryonic and extra-embryonic tissues (Winkel and Pedersen, 1988).

The outer cells of the ICM delaminate to produce a layer of cells, one cell thick, that is called the hypoblast (Rugh, 1968, Theiler, 1972). The other internal cells of the ICM give rise to the epiblast from which all embryonic tissues are formed. This has been demonstrated by reconstituting trophectoderm of one glucose phosphate isomerase variant with the ICM of another variant and allowing the embryos to develop. This study revealed that the entire foetus is derived from the ICM (Papaioannou, 1982). In rodents, after implantation, the epiblast becomes suspended on a cylindrical column of tissue which protrudes into the expanding blastocyst cavity. This structure is referred to as the egg cylinder.

1.2.2 Gastrulation and the generation of the body plan

It is now, at about six days of gestation, that the embryo begins to gastrulate and this will terminate about three and a half days later. Gastrulation is a morphological transformation whereby the multilayered larval form is created, by rearrangement of tissues, from an earlier monolayered form. It is during gastrulation that the body plan of the animal is laid down and the anteroposterior axis of the embryo first becomes visible. It is felt that the anteroposterior axis is specified prior to gastrulation and is a consequence of the orientation of the implanting embryo with respect to its uterine environment (Smith 1985). By analogy with chick embryos, it is possible that the anteroposterior polarity is first established in the hypoblast and later induced to the epiblast (Mitrani and Eyal-Giladi, 1981, Mitrani et al., 1983).

In birds and mammals the meso-endodermal lineages delaminate from the epiblast and involute through the primitive streak, which is like a fault line, lying along the anteroposterior axis at the posterior end of the
epiblast. In the mouse, cell division rates rise dramatically in an area just anterior of the centre of the midline, the proliferative zone, and there is a reorientation of cell division at the posterior end (Snow 1977). This marks the beginning of the primitive streak. The new tissue is inserted between the epiblast and the hypoblast. At this stage it would be more appropriate to term the epiblast the ectoderm.

The mouse starts gastrulation with about 500-600 cells and within 24 hours it has produced some 7000-8000 ectodermal and 6000-7000 mesodermal cells (Snow, 1977). Proliferation would seem to be important in murine gastrulation since mitotic inhibitors arrest gastrulation. This is in contrast to *Xenopus* which can continue gastrulation in the presence of such inhibitors (Cooke, 1973). As gastrulation proceeds the meso-endodermal cells extend between the epiblast and hypoblast to form a complete layer that matches the area covered by the ectoderm. Most regions of the epiblast seem to be pluripotent up to the eighth day, although there may be regional differences such that cells from anterior regions prefer to colonise ectoderm while those from distal and posterior regions contribute more to mesoderm (Beddington, 1982, Tam and Beddington, 1987).

Cells that emerge from the anterior part of the primitive streak contribute mainly to the notochord, gut and paraxial mesoderm while cells from the middle region give rise to lateral mesoderm and from the caudal portion to extra-embryonic mesoderm (Snow, 1981, Tam and Beddington, 1987). While the heart has been anatomically described as forming from a group of cells at the most anterior extremity of the embryonic axis it has also been shown that if a portion of the distal tip of the 7 day old egg cylinder is removed heart development is abolished (Rugh, 1968, Theiler, 1972, Snow, 1985).

The notochord is laid down along the mid-line of the embryo and the paraxial mesoderm lies on either side. The early differentiation of the paraxial mesoderm is characterised by the sequential condensation of epithelial balls of mesoderm, called somites, at the cranial end. The mesoderm that gives rise to the somites is termed presomatic and it has been suggested that there exists a prepattern of somitomeres in this tissue that condense into the somites (Tam et al., 1982, Tam, 1988). In more cranial regions the paraxial mesoderm that emerges from the primitive streak forms somitomeres, of which there are thought to be seven, and does not condense into somites (Meier and Tam, 1982). Adjacent to the paraxial mesoderm lies the intermediate mesoderm which is the precursor of the mesonephros (Saxen et al., 1986). It has been suggested, in the rat, sheep and in a primate that the intermediate mesoderm also gives rise to cells of the somatic part of the gonad and this is
thought to also be true in the mouse (Paranko, 1987, Zamboni and Upadhyay, 1982, Yoshinga et al., 1988). The lateral plate mesoderm is arranged in two layers, one that lies beside the hypoblast, the splanchnic mesodermal plate, and one which lies just beneath the ectoderm, the somatic mesodermal plate. The splanchnic plate is associated with the endoderm and contributes to the mesenchymal portions of many visceral organs such as the lung, stomach and gut while the somatic mesoderm contributes to the outer regions of the body such as the limbs (Balinsky, 1981).

It is important to note that there is a general rostral to caudal temporal sequence in maturation. For example, while there are somites condensing in rostral portions of the embryo gastrulation is still occurring more caudally with cells still passing through the primitive streak (Rugh, 1968, Theiler, 1972).

The ectoderm forms both epidermis and neural tissue with the neural plate forming along the midline after the primitive streak has regressed (Rugh, 1968, Theiler, 1972). In avian and Xenopus embryos neural tissue becomes induced by the underlying axial/paraxial mesoderm and this is probably also true of the mouse (Slack 1983). Neurulation involves the neural plate thickening, and the lateral folds rising up and closing to produce the neural epithelium (Martins-Green, 1988). In the mouse the first fusion of the neural folds occurs at the level of the third or fourth somite and proceeds rostrally and caudally (Sakai, 1989). After this initial event there are also a number of other secondary sites of fusion (Sakai, 1989). In the chick it has been suggested that changes in cell shape are important events in the formation and shaping of the neural epithelium (Schoenwolf and Powers, 1987, Smith and Schoenwolf, 1987) and it is felt that the same events may also occur in the mouse (Martins-Green, 1988). Again there is a rostral to caudal maturation, with the neural fold having formed in anterior regions while the plate has not yet been specified in posterior regions (Rugh, 1968, Theiler, 1972).

It seems likely that positional information will be imparted by the mesoderm onto the ectoderm, although no direct experiments have been carried out on the mouse which would address this question. Grafting experiments in avian embryos would seem to suggest that as in amphibia, it is the mesoderm that imparts positional information to the ectoderm (Hara, 1978). This was demonstrated by recombining competent ectoderm with axial mesoderm from different regions along the rostrocaudal axis. The ectoderm was induced to form neural tissue and the regional characteristics of the induced tissue varied and was dependent upon the site of origin of the mesoderm (Hara, 1978).

The neural primordium is also the source of the neural crest cells which
are mesectodermal cells which form many diverse derivatives (Le Douarin, 1982). The neural crest cells migrate from the margin of the neural groove, in cranial regions, and from the dorsal part of the neural tube in the trunk (Smits-van Prooije et al. 1988). In trunk regions these cells form the neurons and glia of the peripheral nervous system, pigment and adrenomedullary cells. The cranial neural crest has a greater plasticity and can also form cartilage and bone in the face (Le Douarin, 1982).

The eight day mouse embryo has an S shape with its dorsal side lying towards the placenta. This is transformed into a C shape embryo by day 9.5 with the ventral surface lying towards the placenta. This turning is initiated at about the level of the eighth somite and from here it spreads both caudally and rostrally (Poleman et al., 1987). The mechanism that lies behind this axial rotation is unknown although it has been suggested that asymmetric mitotic activity in the neural tube may be involved (Poelmann et al. 1987). 9.5 days of gestation also marks the time of the closure of the caudal neuropore. The primitive streak that probably persists to this time is replaced by the caudal eminence or end-bud (Rugh, 1968, Theiler, 1972). This caudal eminence gives rise to notochord, hind gut, caudal somites and the neural cord. Once the caudal neuropore has closed neural material is not laid down as a plate but as a cord. The cavity of the already formed spinal cord extends into the neural cord. This process is termed secondary neurulation (Schoenwolf, 1984).

The somites break down to form the dermatome, myotome and sclerotome from about 9 days onwards, and again this progresses in a rostral to caudal sequence (Rugh, 1968, Theiler, 1972). The dorsolateral portion of the somites produces the dermatome which contributes to the dermis, and the myotome which contributes to the skeletal muscle. The sclerotome, which is derived from the ventral medial part of the somite, condenses around the notochord and forms the pre-vertebrae which become the backbone of the animal. The notochord degenerates as development proceeds. At this stage one can also identify the fore limb buds, the hind limb buds have not yet formed, and many of the major organs are beginning to differentiate (Rugh, 1968, Theiler, 1972). The lung and stomach forming from the foregut, and the hindgut forms caudally (Sorokin, 1965, Bryden et al., 1973, Rugh, 1968, Theiler, 1972). The mesenchymal components of these organs are probably derived from the splanchnic mesodermal plate. Epithelial-mesenchymal interactions are very important in the development of these organs. For example in the lung it is the mesenchyme at the tip of the developing alveoli that induces branching. This mesenchyme will induce branching on other regions of the lung if transplanted (Sorokin, 1965).

Neurogenesis in the mouse occurs between about 9 and 14 days of gestation.
(Wentworth, 1984a, 1984b). It has been noted that there is a general rostral to caudal sequence in the maturation of the neurons in the rodent nervous system (Altman and Bayer, 1984). These authors also describe a general ventral to dorsal gradient in the maturation of the major classes of neurons. The first major class of spinal neurons to be born are the motor neurons of the ventral horn while the inter-neurons of the dorsal horn are among the last born.

At day 15 of gestation, all the major organ systems are established, the animal has formed all 65 somites and looks rodent-like (Rugh, 1968, Theiler, 1972). The remainder of the gestation period which in total lasts for about 20-21 days is concerned with the differentiation of the cell types of each organ system and with the growth of the embryo.

1.3 Strategies Towards a Molecular Understanding of Murine Development

This description of murine embryogenesis has hardly touched upon the mechanisms or molecules that could be involved in the regional specification that is morphologically obvious. Yet clearly one would wish to move beyond such descriptive studies and to begin to define important molecules that are involved in mammalian development. There are a number of approaches that are currently being attempted in the hope of isolating such molecules.

One approach is to try and clone classically identified genes that exhibit interesting phenotypes such as Brachyury, which affects the development of the notochord, or rachiterata, which affects the pre-vertebrae. In the case of Brachyury, long range walking and cloning of chromosomal regions has lead to the identification of the gene and analysis of this gene is providing insight into the formation of mesoderm (Hermann pers. comm.). This approach has the obvious advantage that once the gene is cloned the mutant phenotypes already exist therefore aiding both further genetic and molecular studies. Another approach is to work with molecules that have been shown to be important in other biological processes such as cellular transformation. The large range of growth factors would be examples of this class and molecules related to the fibroblast growth factors and the transforming growth factors such as XTC-MIF and VglI have been shown to have important roles in Xenopus embryogenesis (Slack et al., 1988, Smith, 1989). It has also been recently shown that the proto-oncogene c-kit maps to the classically defined mouse locus, white (Chabot et al., 1988, Geissler et al., 1988). This gene is involved in the development of stem cells in the haemopoetic, neural crest and germ cell lineages.

A number of groups have attempted to create random recessive mutants
through insertional mutagenesis in the hope of inactivating important developmental genes (Gridley et al., 1987). This approach has used DNA micro-injection into fertilised eggs, retroviral infection of early embryo, blastocysts and midgestation embryos. One then breeds the animals to homozygosity and looks for an interesting developmental phenotype. One retrovirally induced mutation that exhibits an embryonic lethal phenotype, \textit{mov} \textit{I3}, has been shown to disrupt the \textit{al} (I) collagen gene (Schnieke et al., 1983).

Embryonic stem (ES) cells are derived from the early embryo and can be maintained in culture as undifferentiated pluripotent stem cell lines. In culture one can manipulate these cells and then reintroduce the cells back into the embryo where they retain the ability to contribute to all the tissue of the animal including the germ line. Thus it is possible to mutate a gene in tissue culture and introduce that mutation into the animal. Using this technology one could produce a range of mutations ranging from large disruptions to subtle changes in the coding sequence.

One approach towards identifying developmentally interesting genes has been to attempt to tag these genes with reporter genes. This can be most effectively achieved using ES cells since transfection of the reporter construct into plates of cells allows one to amass many different integration events. The different transfected cells can be stored and analysed at a later date. The developmental pattern of expression of a given clone of transfected cells is then analysed by reintroducing the modified ES cells into the blastocyst and thereby creating a chimaeric animal. If the expression pattern proves to be interesting the sequences flanking the integrated reporter can be cloned out and the endogenous gene examined. In this strategy the integration of the vector should create a mutation (Rossant and Joyner 1989) which can then be passed through the germ line of the animal and bred to homozygosity.

However, the approach which has yielded the broadest spectrum of developmental control genes is based on conservation in protein and nucleic acid motifs between different species. This involves searching for murine genes that are structurally related to genes which have been identified as being developmentally important in other organisms. One obviously hopes that if a murine cognate is found then this gene will be important in murine embryogenesis. This means, for example, taking \textit{Caenorhabiditis elegans} or \textit{Drosophila melanogaster} genes and looking for cognate genes in the murine genome.
1.4 Regional Specification in Drosophila Embryogenesis

The fruit fly Drosophila melanogaster has been intensively used in genetic research for over 70 years and it is in this organism that we have the best understanding of the genetic basis of development. Because of its short life cycle, it is possible to screen large numbers of embryos for mutations affecting the basic body plan. Many important genes have been identified by mutant screens that selected for defects in the pattern of the cuticle (Nusslein-Volhard and Weischaus, 1980, Weischaus et al., 1984, Jurgens et al., 1984, Nusslein-Volhard et al., 1984). Drosophila is also the organism in which molecular biology has made the biggest impact in the understanding of the events underlying pattern formation. The embryology of Drosophila melanogaster has been described in detail many times before, such as in Campos-Ortega and Hartenstein (1985), as has the molecular genetics of embryonic pattern formation in Drosophila, such as in Akam, (1987) and Ingham, (1988). I will only present an outline of the important events in the specification of Drosophila embryonic pattern.

The Drosophila egg is ellipsoidal in shape and polarised so that the future axes, the anteroposterior, dorsoventral and left-right axes, of the embryo can be distinguished. The early stages of Drosophila embryogenesis are different from that of many animals in that the zygotic nucleus undergoes a series of rapid, synchronous nuclear divisions that are not accompanied by cytokinesis. This leads to the formation of a syncytium, in which 512 nuclei are located in a common cytoplasm. The subdivision into cells is delayed. A small group of nuclei move into the cortical cytoplasm at the posterior pole of the egg and become surrounded by plasma membranes. These cells form the pole cells which later become the germ cells. Most of the remaining nuclei colonise the cortical cytoplasm at the periphery of the egg. They undergo another four divisions at this site until the actual cleavage of the cytoplasm occurs and the plasma membranes separating the nuclei form. This monolayer of cells around the periphery of the egg is called the blastoderm. Many of the signals that progressively establish the regional specification of the Drosophila embryo act at the syncitial blastoderm stage.

1.4.1 The genetics of anteroposterior specification

The first acting genes are maternal genes and it is these genes that are responsible for establishing the polarity of the egg. Genes of the bicoi d (bcd) group are responsible for determining the anterior structures while those of the oskar (osk) group specify posterior regions and the terminal structures rely on the genes of the torso group (Fronhofer et al., 1986, Lehman and Nusslein-Volhard, 1986, Nusslein-Volhard et al., 1987). The
anterior determinant would seem to be the product of the bcd gene which acts in a positive instructive fashion to organise the anterior pattern (Fronhofer et al., 1986,). Contrastingly the gene of the osk group that may be the posterior determinant, the nanos (nos) gene, is felt to have a permissive role in the posterior pattern (Hulskamp et al., 1989, Irish et al., 1989).

These maternal genes interact with the zygotic genome to activate the gap genes which are responsible for the primary subdivision of the embryo. These genes are so named because mutations in these genes cause deletions in particular parts of the body, which creates a gap in the anteroposterior pattern (Nusslein-Volhard and Wieschaus, 1980). There are three principal members of this class, hunchback (hb), Kruppel (Kr) and knirps (kni) (Nusslein-Volhard and Wieschaus, 1980). hb is expressed in two broad domains, one extending from the anterior pole to 50% of the egg length (EL) and the other from the posterior pole to 25% of EL (Tautz, 1988). Kr is expressed in a broad band in the middle of the embryo and kni is expressed both just anterior and posterior of the Kr band (Tautz and Pfieifle, 1989).

In bcd mutant embryos there is no zygotic expression of hb but Kr is expressed more anteriorly than normal and in a broader domain (Tautz, 1988). The Kr domain of expression is also extended in the absence of the posterior determinants (Gaul and Jackie, 1987). These results have been interpreted as illustrating that the genes of the bcd and osk groups act to repress Kr in both anterior and posterior regions. The activation of hb by bcd has been shown to occur in a concentration dependent fashion (Struhl et al., 1989, Driever et al., 1989b). The bcd protein contains a DNA binding domain and it has been shown that there are a number of bcd binding sites upstream of the hb promoter, which exhibit different affinities for this protein (Berleth et al., 1988, Dreiver and Nusslein-Volhard 1989a, Struhl et al., 1989, Dreiver et al. 1989b).

The hb gene besides being zygotically activated in anterior regions also displays a maternal transcript that is located throughout the egg (Tautz et al., 1987). This maternal expression does not normally persist in posterior regions but in mutant embryos of the osk class this expression does persist. It has been shown that ectopic expression of hb in posterior regions acts to suppress abdominal segmentation and therefore acts like mutants from the osk group (Hulskamp et al., 1989, Struhl, 1989a). It is thought that the role of the nos gene is in clearing this posterior expression of hb (Hulskamp et al., 1989, Irish et al., 1989, Struhl, 1989a). This is supported by the observation that hb,nos double mutants develop a normal abdomen (Hulskamp et al., 1989, Irish et al., 1989).
Analysis of the expression of the gap genes in different gap gene mutant embryos provides evidence for bilateral regulatory interactions amongst the gap genes (Jackle et al., 1986). Thus this type of interaction with mutual repression between adjacent domains results in the sharpening and generation of stable boundaries. These interactions are likely to be at the transcriptional level since hb, Kr and kni encode proteins that have motifs characteristic of many transcription factors (Rosenberg et al., 1986, Tautz et al., 1987, Nauber et al., 1988).

The gap genes act to cause position specific expression of the pair-rule genes, which further subdivide the embryo (Frasch and Levine, 1987b). There are eight members of the pair-rule class of genes, all of which show transient expression in seven or eight stripes during cellularisation of the blastoderm (Akam, 1987). Of these genes runt and hairy would seem to be the first to act and to have a major role in setting up the striped pattern of the other pair rule genes (Ingham and Gergen, 1988c). These two genes are initially expressed uniformly but by the beginning of cellularisation they are expressed in two complementary series of seven stripes around the embryo and may act to interpret the positional information established by the maternal and gap genes (Howard, 1988). It is not clear exactly how these patterns are established but there would seem to exist an interaction between the gap genes and between the runt and hairy gene products themselves and this is supported by the analysis of gene expression in mutant embryos (Ingham and Gergen, 1988c). Once established the complementary stripes of runt and hairy serve as the driving force for the refinement of the spatial patterns of the other pair-rule genes.

The pair-rule genes will act to define the domains of expression of the segment polarity genes, such as engrailed (en) and wingless (wg), that are the first evidence of the parasegmental organisation (Martinez-Arias and Lawrence, 1985). The en stripes represent the anterior limit of a parasegment and the wg stripes are localised to the posterior cells of each parasegment (Ingham et al., 1985, Baker, 1987). It has been demonstrated that alternate en stripes require the combined activity of different sets of pair-rule genes. In odd numbered stripes, the en expressing cells are those which express both even-skipped (eve) and paired (prd), while in the even-numbered stripes they are those that express fushi-tarazu (ftz) and odd-paired (Ingham et al., 1988b). In contrast to en, wg expression appears to be repressed by the eve and ftz proteins (Ingham et al., 1988b). At the end of the blastoderm stage the eve and ftz bands narrow so that they are separated by a single cell, and it is this cell that expresses wg.

It is also at this stage that each parasegment primordium is programmed
to follow a unique differentiation pathway and this depends on the selective action of the homeotic genes of the Antennapedia (ANT-C) and Bithorax (BX-C) complexes (Akam, 1987). The term homeotic comes from the mutant phenotypes of these genes which cause one segment type to be transformed to that of another homologous structure (reviewed in Gehring and Hiromi, 1986). The homeotic genes are first transcribed at the blastoderm stage at the same time as the pair rule genes and are affected by the maternally acting genes and gap genes, which cause broad regional differences in the expression of the homeotic genes (Harding and Levine, 1988). For example the P1 promoter of the Antennapedia (Antp) gene depends absolutely upon Kr while the p2 promoter is acted upon by a combination of osk, hb and ftz (Irish et al., 1989). This pattern is then refined by the pair-rule genes themselves. For example it has been demonstrated that ftz, which is involved in the correct regulation of en, is also necessary for the expression of Sex combs reduced (Scr), Antennapedia (Antp) and Ultrabithorax (Ubx) (Ingham and Martinez-Arias, 1986). It is clear that the homeotic genes act to modulate their own transcription and it has been demonstrated that the Deformed (Dfd) gene exhibits autoactivation (Kuziora and McGinnis 1988) and it has also been shown that the more posteriorly expressed homeotic genes will act to repress the more anteriorly expressed genes, such as the repression of (Ubx) by Abdominal-B (Abd-B) and abdominal-A (abd-A) (Hafen et al. 1984, Harding et al. 1985, Struhl and White 1985).

These interactions eventually result in defined domains of homeotic gene expression which allow these genes to specify parasegment character. For example Scr, Antp, and Ubx exhibit enhanced levels of expression in the primordia of parasegments 2, 4 and 6 respectively and these are reflected in the mutant phenotype of these genes (Martinez-Arias et al., 1987, Carroll et al., 1986a, Akam and Martinez-Arias, 1985). The Scr mutants show a transformation of the labial segment, which is partly derived from parasegment 2, whereas Antp mutants have a transformation of the second thoracic segment, which has a contribution from parasegment 4 (Kaufman et al., 1980, Wakimoto and Kaufman, 1981). Similarly the first abdominal segment which is derived from parasegment 6 is transformed in a Ubx mutant (Sanchez-Herrero et al., 1985).

1.4.2 The genetics of dorsoventral specification

The dorsoventral polarity is also thought to be established through a hierarchy of interactions that start with maternally encoded gene products. The first identified maternally acting gene was dorsal (Nusslein-Volhard, 1979). This gene exhibits an extreme phenotype in which the embryo becomes
completely dorsalised (Santamaria and Nusslein-Volhard, 1983). Of the other maternal genes that have been identified Toll seems to be acting early in the generation of dorsoventral polarity. Rescue of Toll mutant embryos can be effected with cytoplasm taken from any part of the wild type egg (Anderson et al., 1985). Yet in contrast to other mutants of this class, such as dorsal, the site of injection defines the position at which the ventral most structure will form. As with dorsal the Toll RNA, is not localised but is uniformly distributed in the early embryo (Hashimoto et al. 1988). The Toll protein is thought to become modified or localised in the ventral region of the embryo, and it is thought that this may be dependent upon Toll activity itself (Levine 1988a). Levine (1988a) suggests that it is the Toll product which is directly or indirectly involved in the localisation of the dorsal activity to ventral regions of the embryo and that it is this activity that may play a key role in the differentiation of dorsoventral pattern elements.

Again, as with the anteroposterior network, it seems that the zygotic genes act to interpret and further subdivide the maternal dorsoventral information. A number of such genes have been identified. For example Zerknult (zen) and decapentaplegic (dpp) are two such genes which have null mutants which cause the loss of dorsal derivatives and exhibit a ventralised phenotype (Wakimoto et al., 1984, Irish and Gelbart, 1987). Contrastingly the elimination of mesoderm is the observed phenotype of mutants in the twist and snail genes, while the single-minded mutants wipe out both the neuronal and non-neuronal derivatives of the ventral most ectoderm (Simpson, 1983, Thomas et al., 1988). The patterns of expression of some of these genes is initially quite broad and becomes defined as gastrulation proceeds (Ingham 1988a). A number of these genes contain regions that would encode DNA binding domains. zen contains a homeo domain and snail has a zinc finger motif (Rushlow et al., 1987, Boulay et al., 1987). Therefore these genes may exert their effects through transcriptional regulation. Other zygotically acting genes such as dpp display identity to growth factor like molecules and could therefore be involved in cell-cell signalling (Padgett et al., 1987). Thus while there is much less molecular detail about the sequence of events in the establishment of dorsoventral polarity it is obvious that it shows parallels with the establishment of anteroposterior polarity and that it is separate from the anteroposterior polarity.

1.4.6 Neurogenesis

Many of the segmentation and homeotic genes are not only expressed during early embryogenesis but are also expressed at later times in larval and imaginal structures (e.g. Dfd and Scr, Martinez-Arias et al., 1987). The
embryonic nervous system, in particular, shows high levels of expression of many of these genes. The segmentation genes eve, en, and ftz show expression in a specific subset of neurons in every segment of the developing CNS (Frasch et al., 1987a, Dinardo et al., 1985, Carroll and Scott, 1985). These genes seem to be involved in neuronal determination and it has been shown that ftz and eve are components of the mechanism controlling cell fate during neuronal development (Doe et al., 1988b, Doe et al., 1988c). It was also demonstrated that the segmentation genes also interact with each other in the developing nervous system, although not in the same fashion as was observed in the blastoderm (Doe et al. 1988c). In the central nervous system loss of ftz affects the expression of eve while the opposite situation is found in the blastoderm. This would suggest that these genes are responding to different cues in the nervous system as as opposed to those in the blastoderm.

The homeotic genes are expressed at their highest levels in embryonic nervous system (Levine et al., 1983, Akam and Matinez-Arias, 1985, Chadwick and McGinnis, 1987, Kuriowa et al., 1985, Martinez-Arias et al., 1987). In this system they also appear to be involved in conferring positional specification. Thus they are not expressed in every segment of the nervous system but exhibit higher levels of expression in specific parasegments and some genes are expressed at lower levels in the surrounding parasegments. Mutations in these genes affect those regions that exhibit the high levels of expression of the genes. As has been described for the early embryo the homeotic genes in the central nervous system still exhibit cross regulatory interactions, such as the repression of anterior gene expression by posterior genes.

This summary of *Drosophila* development clearly points out the use of a hierarchy of decisions, through gene activation, to specify the pattern of the *Drosophila* body. Genes for the anteroposterior axis initially act to subdivide the embryo till the parasegments are defined and then finally through the action of the homeotic genes these metameric units are given an identity. The hierarchical action of genes is also evident in the establishment of the dorsoventral axis.

1.5 The Homeobox

As is evident from the discussion above there are many interesting developmental genes in *Drosophila*, that have been identified by classical genetics. One particularly interesting class of genes are the homeotic genes which can cause the transformation of one body segment into another. These
genes act to interpret the anteroposterior positional information and confer identity on the parasegment primordia. There was obviously a great deal of interest in cloning these genes and analysing their sequences. The molecular analysis of these genes revealed a common sequence that was present in two of the homeotic genes, \textit{Antp} and \textit{Ubx}, and in one segmentation gene (\textit{ftz}). This common element, which was termed the homeobox because it was discovered in homeotic genes, is a conserved 183 base pair sequence of DNA that is present within the protein coding regions of these genes (McGinnis et al., 1984, Scott and Weiner, 1984). Subsequent analysis has shown that the homeobox is present in all the homeotic genes of the ANT-C and BX-C complexes and in genes such as \textit{bcd} and \textit{en} that have been discussed above.

Sequence analysis has shown that the protein domain that is encoded by the homeobox, the homeodomain, is structurally related to the prokaryotic DNA binding domains that have the helix-turn-helix motif (Gehring, 1987). It has been shown that the homeodomain is a DNA binding domain and homeodomain proteins are transcriptional regulators (Levine and Hoey, 1988b). Further molecular characterisation of genes from \textit{Drosophila}, vertebrates, and other organisms has shown that there are actually quite a range of homeobox type sequences. The major classes would include those related to one of \textit{Antp}, \textit{en}, \textit{eve} or \textit{caudal (cad)} and a new class that contain the POU-homeodomain. The POU homeodomain has been found in three mammalian transcription factors Pit-1, Oct-1 and Oct-2 and one \textit{Caenorhabditis} gene \textit{unc-86}, their homeodomains while being related to the \textit{Drosophila} homeodomains are actually more similar to each other (Bodner et al., 1988, Ingraham et al., 1988, Sturm et al., 1988, Clerc et al., 1988, Finney et al., 1988). This has been termed the POU homeo domain for the Pit-1, Oct-1 and Oct-2 and unc-86 genes (Herr et al., 1988). These genes are also related by virtue of a second region the POU-domain, that is about 75 amino acids in length. These different classes of homeodomains differ from each other at many points over the 61 amino acids but can be grouped together by considering their amino acid changes in the most conserved region of the homeodomain, the helix-turn-helix region. This thesis will largely concern itself with homeoboxes/homeodomains of the \textit{Antp} type and unless otherwise stated the term homeobox/homeodomain shall refer to those of the \textit{Antp} type.

An initial study of the phylogenetic distribution of homeoboxes suggested that the presence of this sequence may correlate with a segmented body plan (McGinnis, 1985). Yet it was subsequently shown that homeoboxes were evolutionarily more widespread than was previously appreciated with homeoboxes being found in organisms of widely separate phyla from non-segmented
invertebrates to mammals (Holland and Hogan, 1986, Costa et al., 1988). Of the
genesis in which the homeobox was first discovered, only ftz is a segmentation
gene while the others are responsible for anteroposterior specification
(Slack, 1984). Later studies of other Drosophila homeotic genes uncovered
patterns of expression that suggested roles in positional information
independent of the process of segmentation (Hoey et al., 1986). Thus the
specification of positional values along the anteroposterior axis of an animal
may be a general role for homeotic genes.

There have been many homeobox genes cloned from the mouse genome, and
there are at least 27 such genes. As with Drosophila these genes are found in
clusters in the genome and exhibit temporally and spatially restricted
patterns of expression during development. Aspects of the structure and
expression of these genes will be dealt with later in this thesis. The
expression patterns of these genes would be consistent with them having a role
in important developmental processes. However it is possible that the
conservation of the homeobox sequence between mouse and Drosophila represents
evolutionary/developmental economy, in that if one has a useful protein
structure one keeps employing that structure although not necessarily for the
same purpose. Alternatively conservation could represent true conservation of
function. In the latter case one would predict that the murine homeobox genes
are also acting to interpret anteroposterior positional information. If this
were true and given that Drosophila and mammals represent the peaks of the two
major coelomate lineages, the protostomes and the deuto stomes respectively,
one would be implying that there has existed common biochemical elements for
the interpretation of anteroposterior positional information since before the
split between these two major lineages.
CHAPTER TWO - MATERIALS AND METHODS

2.1 Buffers and Stock Solutions

Alkaline SDS 0.2M NaOH, 1% SDS.

Blue juice (X10) 50% glycerol, 20mM Tris.Cl, 20mM EDTA (pH 8.2), 0.1% Bromophenol blue.

Denaturing solution 0.5M NaOH, 1.5M NaCl.

Denhardt's solution (100X) 2% (w/v) BSA fraction V, 2% (w/v) Ficoll 400, 2% (w/v) polyvinylpyrrolidone. Filter sterilised and stored at -20°C.

DEPC-treated water 0.1% (v/v) DEPC was added and then autoclaved.

DTT, 1M 1M DTT in 10mM NaOAc (pH 5.2).
Sterilised by filtration. Stored at -20°C.

EDTA, 500mM pH adjusted to 8.0. Autoclaved. For RNA work this solution was also treated with DEPC prior to autoclaving.

Elution buffer 10mM Tris.Cl (pH 7.6), 1mM EDTA (pH 8.0), 0.1% SDS.

Hybridisation buffers

In situ 50% deionised formamide, 0.3M NaCl, 20mM Tris.Cl (pH 8.0), 5mM EDTA, 10mM NaPO₄ (pH 8.0), 10% Dextran sulphate, 1X Denhardt's, 0.5mg/ml yeast RNA.

Riboprobe pre-hyb 50% deionised formamide, 5X SSC, 5X Denhardt's, 50mM NaPB (pH 6.8), 250ug/ml
Salmon sperm DNA, 100µg/ml tRNA, 10µg/ml poly A, poly C, 0.2% SDS.

Riboprobe hyb 50% deionised formamide, 5X SSC, 1X Denhardts, 20mM NaPB (pH 6.8), 100µg/ml Salmon sperm DNA, 100µg/ml tRNA, 10µg/ml poly A, poly C, 0.2% SDS, 10% Dextran sulphate.

Ligase buffer (X5) 250mM Tris.Cl (pH 7.6), 50mM MgCl₂, 25% (w/v) polyethylene glycol 8000, 5mM DTT

Lysis buffer (mini-prep) 8% sucrose, 2% TritonX-100, 50mM EDTA (pH 8.0), 50mM Tris.Cl (pH 8.0).

Lysis buffer (maxi-prep) 50mM glucose, 25mM Tris.Cl (pH 8.0), 10mM EDTA (pH 8.0).

MOPS (X20) 0.4M morpholinopropanesulfonic acid (MOPS) (pH 7), 50mM NaOAc, 10mM EDTA (pH 8). Filter sterilised and stored at 4°C.

Neutralising solution(southern) (colony screen) 0.5M Tris.Cl (pH 7.6), 3.0M NaCl.

PBS (X10) 1.3M NaCl, 70mM Na₂HPO₄, 30mM NaH₂PO₄. pH was adjusted to 7 and autoclaved. For RNA work it was treated with 0.1% DEPC and then autoclaved.

Probe hydrolysis buffer 80mM NaHCO₃, 120mM Na₂CO₃, 10mM DTT (pH 10.2). Filter sterilised and stored at -20°C.

Probe neutralising buffer 0.2M NaOAc (pH 6.0), 1% (v/v) acetic acid, 10mM DTT. Filter sterilised and stored at -20°C.
### Restriction enzyme buffers

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low salt (X10)</td>
<td>100mM Tris.Cl (pH 7.6), 100mM MgCl₂, 10mM DTT.</td>
</tr>
<tr>
<td>Medium salt (X10)</td>
<td>500mM NaCl, 100mM Tris.Cl (pH 7.6), 100mM MgCl₂, 10mM DTT.</td>
</tr>
<tr>
<td>High salt (X10)</td>
<td>1M NaCl, 500mM Tris.Cl (pH 7.6), 100mM MgCl₂, 10mM DTT.</td>
</tr>
<tr>
<td>Very high salt (X10)</td>
<td>1.5M NaCl, 500mM Tris.Cl (pH 7.6), 100mM MgCl₂, 10mM DTT.</td>
</tr>
<tr>
<td>EcoRI buffer (X10)</td>
<td>1M NaCl, 1M Tris.Cl (pH 7.6), 100mM MgCl₂.</td>
</tr>
<tr>
<td>Smal buffer (X10)</td>
<td>200mM KCl, 100mM Tris.Cl (pH 8.0), 100mM MgCl₂, 10mM DTT.</td>
</tr>
</tbody>
</table>

All restriction enzyme buffers were filtered sterilised and stored at -20°C.

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saline (X10)</td>
<td>8.3% (w/v) NaCl. Treated with 0.1% DEPC and then autoclaved.</td>
</tr>
<tr>
<td>Sodium acetate, 3M</td>
<td>pH adjusted to 5.2 with glacial acetic acid. Autoclaved. For RNA treated with 0.1% (v/v) DEPC then autoclaved.</td>
</tr>
<tr>
<td>Sodium phosphate buffer (NaPB)</td>
<td>1M Na₂HPO₄ mixed with 1M NaH₂PO₄ until pH 6.8. Autoclaved.</td>
</tr>
<tr>
<td>SSC (X20)</td>
<td>3M NaCl, 0.3M Na citrate pH adjusted to 7. For RNA treated with 0.1% DEPC and autoclaved.</td>
</tr>
<tr>
<td>Transcription buffer (X10)</td>
<td>400mM Tris.Cl (pH 7.6), 60mM MgCl₂, 20mM spermidine, 200mM NaCl.</td>
</tr>
<tr>
<td>Tris.Cl, 2M</td>
<td>pH adjusted to 7.6 and 8.0. For RNA work this solution was made up with DEPC treated water.</td>
</tr>
<tr>
<td>Tris-acetate (TAE) X10</td>
<td>40mM Tris-acetate, 1mM EDTA (pH 8.2).</td>
</tr>
</tbody>
</table>
Tris-borate (TBE) X10 900mM Tris base, 900mM boric acid, 20mM EDTA (pH 8.2).

TE 10mM Tris.Cl (appropriate pH), 1mM EDTA (pH 8.0). Autoclaved.

2.2 Bacterial Media, and Antibiotics

Media
L-agar 1% (w/v) bacto tryptone, 0.5% (w/v) bacto yeast extract, 0.5% (w/v) NaCl 1.5% (w/v) bacto-agar.

L-broth As L-agar but omit bacto agar.

Antibiotics
Ampicillin Dissolved at 50mg/ml in sterile water filtered sterilised and stored at -20°C. Added to media and agar to a final concentration of 50-100ug/ml.

Tetracycline Dissolved at 20mg/ml in 50% (v/v) ethanol, filter sterilised and stored at -20°C. Added to media and agar at a final concentration at 20ug/ml.

2.3 Animals
The mice used in this study were of the CBA inbred strain of mus musculus, maintained by NIMR biological services division. The age of the embryos was determined from the day on which the vaginal plug was found. The day on which the plug was found was called day 0 of pregnancy. At noon the next day the embryos are 1.5 day p.c.

2.4 Cloning and DNA Methods

2.4.1 Bacterial strain
The only bacterial strain used in this work was the JM 109 strain of Escherichia Coli K12. The strain is RecA1, endA1, thi, HsdR17, supE44, relA1,
lambda\textsuperscript{-}, del(lac-proAB), (F\textsuperscript{'} , traD36, proAB, lacI\textsuperscript{97}-del M15).

2.4.2 The production of cells competent for Transformation

This method is basically that of Maniatis et al (1982).

Throughout all steps in this procedure solutions and centrifuge bottles were sterile and chilled on ice. Bacterial cells were streaked out, for single colonies, on an L-agar plate and incubated at 37°C. After overnight incubation a single colony was used to inoculate two 5ml l-broth aliquots which were again grown overnight. Each of two flasks containing 500ml of L-broth were inoculated with one of the 5ml overnight cultures. The flasks were then incubated in a shaker incubator at 37°C until the $O.D_{550}$ of the culture reached 0.45 - 0.55. When the cells had reached this point the cultures were chilled on ice, to restrict further growth. The cells were then pelleted by centrifugation at 5k for 5-8 minutes. The supernatant was poured off and each pellet was gently resuspended in 125ml of 0.1M MgCl\textsubscript{2}. The two MgCl\textsubscript{2} suspensions were combined into a single centrifuge bottle and again pelleted as above and then resuspended in 250ml 0.1M CaCl\textsubscript{2} solution. They were left for 20 minutes on ice and then again pelleted. Again the supernatant was discarded and the cells were resuspended in 25ml of a CaCl\textsubscript{2}/glycerol solution. This solution consisted of 21.5ml of 0.1M CaCl\textsubscript{2} and 3.5ml of glycerol. 250ul aliquots of these cells were quick frozen and stored at -70°C.

2.4.3 Bacterial transformation

An aliquot of CaCl\textsubscript{2} treated bacterial cells was thawed slowly on ice. Typically not more than 1 - 0.1ug of DNA, be it a closed plasmid or the product of a ligation reaction, was used to transform cells. The DNA which was dissolved in 50ul of TE was added to 100ul of cells. These were incubated on ice for 20-30 minutes and then heat shocked at 42°C for 2 minutes. This mix was inoculated into 5ml of L-broth and incubated, with shaking, at 37°C for 60-90 minutes. Various amounts of the 5ml culture, ranging from 100ul to 500ul to the remainder were plated on L-agar plates containing the appropriate selectable marker. Plates were then incubated overnight at 37°C. Single colonies were picked the next day for subsequent processing, as required.

2.4.4 Standard DNA methods

Routine procedures involved in DNA manipulation (e.g. phenol: chloroform extraction and ethanol precipitation) were performed as described in Maniatis et al (1982). DNA samples were stored at -20°C.
2.4.5 Restriction digests

All restriction enzymes were used according to the manufacturers instructions. In this lab one of seven restriction enzyme buffers was used depending on the enzyme. These buffers are low (L), medium (M), high (H) and very high (VH) salt buffers. There are also specific buffers for EcoR1 (R1), and Smal (Sma).

2.4.6 Agarose gel electrophoresis

Restriction digests, or other DNA samples, were analysed by electrophoresis through agarose gels. The percentage of each gel varied depending on the size of DNA fragments that were to be analysed. The gels consist of agarose dissolved in either 1X TBE or 1X TAE and run in the same buffer. TAE was used as a gel running buffer when a slow run was required i.e. for southern blotting or precise mapping. The restriction digest or DNA solution was added to the gel in blue juice. DNA fragments were visualised by staining with the fluorescent dye, ethidium bromide which fluoresces strongly under ultra-violet light. This dye was either included in the gel at 0.5ug/ml or after electrophoresis the gel was bathed in a solution containing the ethidium bromide at the same concentration. Fragment sizes were determined by running DNA markers of known molecular weight. The set of fragments that were commonly used can be obtained from BRL as the 1kb ladder.

2.4.7 Isolation of DNA fragments from gels

If a fragment was to be isolated from a gel the gel was stained after electrophoresis. The fragment of appropriate size was cut out of the gel using a sharp scalpel. The gel slice was placed in a dialysis bag that was clipped at one end with a dialysis clip and contained 0.5X TBE. The slice was allowed to fall to the bottom of the bag and the rest of the buffer was removed. This left just enough buffer to cover the slice. The bag was immersed in a shallow layer of 0.5X TBE in an electrophoresis tank and electrophoresed at 100 volts for 2-3 hours. The polarity of the current was then reversed for 40 seconds. The fluid from the bag was removed and the bag was then washed with a small volume of 0.5X TBE. The wash was added to the bag fluid. The 0.5X TBE, that now contains the DNA, was passed over a column of packed siliconised glass wool. The DNA was then cleaned by phenol: chloroform extraction and ethanol precipitation.

2.4.8 Sub-cloning and ligations

Fragments of interest were subcloned in pKS vectors (Stratagene) for
subsequent amplification and to allow in vitro transcription.

Vector DNA was digested with the appropriate enzymes, the digestions checked on a gel and then the DNA was phenol: chloroform extracted and ethanol precipitated. DNA fragments to be subcloned were isolated from the gel as described above. "Sticky end" ligations generally used 1ug of vector DNA and 1 to 3 molar equivalents of insert DNA. The ligation reaction volume was 100ul and contained IX ligase buffer, 1mM ATP and 1u of T4 DNA ligase. The ligation was carried out for six hours at room temperature and the products were used to transform bacteria.

2.4.9 Isolation of plasmid DNA - Mini-prep

A single colony was inoculated into a 5ml L-broth culture, containing the appropriate selection, and grown overnight in a shaking incubator at 37°C. The cells were pelleted by centrifugation at 3k for 10 minutes. The pellet was resuspended in 700ul of lysis buffer. This suspension was transferred to a micro-centrifuge tube and to this 60ul of a lysozyme solution, 10mg/ml in 10mM Tris (ph8.0), was added. These tubes were incubated at room temperature for 10 minutes. The tubes were then placed in a vigorously boiling water bath for 40 seconds. After, they were spun immediately in a microcentrifuge for 7 minutes. The glutinous pellets were fished out with a toothpick and isopropanol was added to the top of the tube. The tubes were vortexed and then spun for 15 minutes. The alcohol was sucked off and the pellet was redissolved in 200ul of TE and then phenol: chloroform extracted and ethanol precipitated. The pellet was redissolved in 20ul of TE.

2.4.10 Isolation of plasmid DNA - Maxi-prep

Large scale isolation of plasmid DNA was used for high purity preparations.

A 5ml overnight L-broth culture, which had been inoculated with a single colony, was added to 500ml of L-broth and grown overnight with shaking at 37°C. In both instances the L-broth contained the appropriate selection for the plasmid. This 500 ml culture was centrifuged at 8k for 10 minutes and the supernatant discarded. The pellet was resuspended in 50ml of lysis buffer. 2.5ml of 40mg/ml lysozyme in lysis buffer was added and incubated at room temperature for 10 minutes. This solution was then mixed with 100ml of alkaline SDS and left on ice for 5 minutes. 75ml of 3M KOAc was added to the preparations, mixed and incubated on ice for 20 minutes. Following centrifugation, at 8K for 10 minutes, the supernatant was poured through a cheese cloth filter. This flow through was mixed with 0.6 volume of
isopropanol and put at -20°C for half an hour. The DNA was recovered by centrifugation at 8K for 10 minutes. The DNA pellet was redissolved in 15ml TE and then ethanol precipitated. This pellet was redissolved in 2ml of TE and the volume then remeasured. To this solution CsCl₂ was added to a final concentration of 1.15g/ml. The final total volume was made up to 5ml with CsCl₂, with a refractive index of 1.393, and also included 300ul of EtBr. This tube was heat sealed. The sample was centrifuged over-night at 55K in a Beckman vti65.2 rotor. The banded plasmid was taken off the gradient and transferred to a new ultra-centrifuge tube. To this was added 75ul of EtBr and the volume was made up to 5ml with the above CsCl₂ solution. The tube was heat sealed and spun at 55k for 6 hours. The band was harvested and the EtBr was removed by isoamyl extraction. The volume was expanded to 15ml with TE and the DNA was ethanol precipitated. The pellet was redissolved in 1ml of TE and the OD₂₅₀ was determined. The DNA was then re-ethanol precipitated and finally dissolved in TE at a concentration of 1ug/ul.

2.4.11 Southern blot

Southern hybridisation was employed in the characterisation of cloned DNA sequences. Restriction digests of DNA were run on agarose gels, as described above. Gels were soaked in denaturing solution twice for thirty minutes and then in neutralising solution, again twice for thirty minutes. Gels were placed on a sheet of Whatman 3mm paper, that was pre-soaked in 20X SSC, which was supported over, but in contact with, a reservoir of 20X SSC (Maniatis et al 1982). A sheet of genescreen, pre-wetted in water and 2X SSC, was placed on top of the gel, followed by two pieces of whatman 3mm paper, again pre-wetted in 2X SSC. On top of this was a stack of level tissues and placed on top of the tissues was a 1kg weight. Transfer was allowed to proceed for 30 minutes for DNA fragments at high concentration and overnight for fragments at low concentration. After transfer the lanes of the gel were marked and the filter rinsed briefly in 6X SSC. The filter was then baked at 80°C for two hours. The dry filter was now ready to be probed which was invariably with a riboprobe. Filters were usually pre-hybridised for 1 to 4 hours in riboprobe pre-hybridisation buffer at the hybridisation temperature. Filters were incubated overnight at 65°C for riboprobes in hybridisation buffer.

2.4.12 Colony screen

This method was used to identified true recombinants in the antibiotic resistant progeny of a bacterial transformation. After the bacterial colonies have grown to a reasonably small diameter, i.e. 0.1-0.2mm, the plate is
removed from the incubator. A piece of dry nitro-cellulose filter is carefully placed on top of the plate. The filter is left in place for a couple of minutes and then peeled off. The filter is placed on whatman 3mm filter paper with the adherent bacterial colonies face up. A second identical filter is then placed directly on top of the first filter and flat pressure is applied. Both filters, while still stuck together, are marked by making needle holes. This enables orientation to be determined afterwards. The two filters are separated and both transferred to new L-agar plates, with the appropriate selection. These plates are grown overnight at 37°C. There are now two copies of the original plate, each on nitro-cellulose filters. One of these filters is kept as a master copy.

The other filter is placed for 5 minutes on filter paper soaked in 10% SDS, then 5 minutes on filter paper soaked in denaturing solution and finally for 5 minutes on filter paper soaked in neutralising solution. The filter is then washed in a solution of 2X SSC, 0.1% SDS. The filter is then baked at 80°C for two hours. The filter was then probed with a riboprobe, as appropriate. The filters are pre-hybridised and hybridised as above. After probing the positive colonies are identified by comparing the signal on the autoradiograph to the master plate, using the marks to align the two.

2.5 RNA Methods

2.5.1 In vitro Transcription

Riboprobes were synthesised essentially as described by Melton et al (1984) using either \(^{32}\)P or \(^{35}\)S labelled TP, as appropriate. The reactions were carried out in 1X transcription buffer and 10mM DTT. The other three ribonucleotides were used at a final concentration of 0.75mM and 25 units of RNasin was included in each 20ul reaction. Either 20 units of T7 or T3 or 10 units of SP6 RNA polymerases were used, as required, to transcribe from lug of linearised DNA template. The transcriptions were at 37°C for 60-90 minutes. Afterwards the DNA template was removed using DNase at a final concentration of 0.025 g per ul and incubating at 37°C for a further 10 minutes. At this step t-RNA, at 1.025ug/ul, was added. The reaction was stopped by the addition of a half volume of stop buffer containing 250mM EDTA, 3mM NaOAc. Unincorporated nucleotides were removed by column chromatography over A1.5m agarose beads. The column was washed in ellution buffer and the excluded peak was pooled.

\(^{32}\)P labelled probes for southern or colony hybridisations were used without further treatment.
$^{35}$S labelled probes for in situ hybridisation were reduced to an average size of 50-150 bases by limited alkaline hydrolysis (Cox et al., 1984). The peak fraction was alcohol precipitated and the pellet then redissolved in 50μl 10mM DTT. To this an equal volume of hydrolysis buffer was added and the probe incubated at 60°C for a time period that was dependent upon the probe length. The time required to reduce a probe to an average size of 100 bases was calculated from the following formula: time in minutes = (original transcript size in kb - 0.1)/(original transcript size in kb x 0.011) (Cox et al., 1984). After hydrolysis an equal volume of neutralising buffer was added and the probe was ethanol precipitated. The pellet was redissolved in 10mM DTT at 1x10$^6$ d.p.m/μl. The hydrolysed product of this reaction was analysed by electrophoresis through an agarose formaldehyde gel followed by transfer to genescreen and then visualisation by autoradiography.

2.5.2 Agarose-formaldehyde gels

RNA is electrophoresed through denaturing agarose-formaldehyde gels. Agarose was melted in water and cooled to 55°C before the addition of 1X MOPS and formaldehyde to 2.1M, from a 12.33M stock. RNA samples were prepared in 1X MOPS, 2.1M formaldehyde, 50% formamide and 1X blue juice and were heated at 65°C for ten minutes. The samples were then cooled on ice for 15 minutes and then loaded into the wells. The gel was run in 1X MOPS at 60-70V for 4-6 hours.

2.5.3 Northern blot

After electrophoresis the agarose-formaldehyde gel was soaked in 50mM NaOH, 0.1M NaCl then in 0.1M Tris.Cl (pH 7.6) and finally in 2X SSC in each case for 20 minutes. The RNA was transferred to genescreen soaked in 2X SSC as described above by blotting in 20X SSC overnight. The lanes were marked and the filter was briefly rinsed in 6X SSC. The wet filter was placed in Saran wrap and exposed to U.V. light for 5 minutes at 600uWatts/cm$^2$. The filter was then baked for two hours at 80°C.

2.5.4 In situ hybridisation

This technique was used to visualise gene transcripts in situ in the embryo.

Microscope slides were cleaned by passing them through 10% HCl in 70% ethanol then water and finally 95% ethanol. The slides were dried in an oven for 15 minutes at 150°C. The dry slides were then placed in a 2% solution of TESPA in acetone for 10 minutes. The slides were then washed twice in acetone
and once in water. They were dried at 50°C.

Embryos were fixed overnight at 4°C in 4% (w/v) paraformaldehyde. They were then washed for 30 minutes, on ice, firstly in PBS and then in saline. The samples were then placed in ethanol: saline (50:50) for 15 minutes followed by 70% ethanol for 2 times 15 minutes. The embryos were dehydrated by passing them through 80%, 95%, 100% and 100% ethanol for 30 minutes at each step and then put in toluene twice for thirty minutes. The toluene was replaced with paraffin wax and the embryos kept at 60°C. The wax was changed three times at twenty minute intervals. The embryos were transferred to a plastic mould containing wax and allowed to cool. The embedded embryos were stored at 4°C.

6 um sections were cut on a microtome and the sections were lowered onto to a 50°C water bath. Creases were allowed to unfold for 2 minutes before the sections were lifted onto subbed slides. The excess water was drained and the slides were placed on 37°C hotplate for 1 hour. The slides were then allowed to dry at room temperature overnight in a box containing desicant. The sections were either used now or stored at 4°C.

Where possible in the following steps all solutions were DEPC treated and diluted from stock into DEPC treated water.

The sections were dewaxed in xylene, twice for 10 minutes. They were then rehydrated by being passed through an alcohol series i.e. 100%, 100%, 95%, 80%, 70%, 50% and 30% ethanol. The slides were washed in saline and then PBS for 5 minutes each and then fixed in 4% (w/v) paraformaldehyde for 20 minutes. They were washed twice for 5 minutes in PBS. The slides were placed horizontal with the sections facing upwards and a proteinase K solution (20ug/ml) was overlain. The proteinase K digestion was allowed to go for 7 minutes before the slides were washed in PBS for 5 minutes. The sections were refixed in 4% (w/v) paraformaldehyde for 5 minutes and then acetylated in fresh 0.25% (v/v) acetic anhydride in 0.1M triethanolamine.HCl (pH 8.0) for 10 minutes. The slides were washed for 5 minutes, firstly in PBS and then in saline. The sections were dehydrated by passing them through an alcohol series (as above). They were air dried and used for hybridisation.

The probes used in in situ hybridisation were in vitro transcribed from linearised DNA and were labelled with $^{35}$UTP. The probe was diluted ten fold in hybridisation buffer to give a final activity of $1 \times 10^5$ d.p.m./ul. About 15ul of hybridisation mix was applied directly to the sections, spread with a strip of parafilm and a clean coverslip was lowered on top. The sections were hybridised for 14-16 hours at 55°C in a box saturated with 50% formamide and 1.25X SSC. After hybridisation the coverslips were dislodged by
soaking in a solution of 5X SSC, 10mM DTT for 20 to 30 minutes. The slides were then washed in 50% formamide, 2X SSC, 0.1M DTT at 65°C for 30 minutes. This was followed by two 10 minutes washes in 0.5M NaCl, 10 mM Tris, 5mM EDTA at 37°C and then a 30 minutes incubation, at the same temperature and in the same buffer, with RNase A at 20ug/ml. After the RNase treatment the slides were washed in the same buffer for 15 minutes before being rewashed in 50% formamide, 2X SSC and 10mM DTT for 30 minutes. Slides were finally washed in 2x SSC and then 0.1X SSC for 15 minutes each and dehydrated through an alcohol series: 30% ethanol, 0.25M NH₄OAc; 50% ethanol, 0.25M NH₄OAc; 70% ethanol, 0.25M NH₄OAc, 80% ethanol, 0.25M NH₄OAc, 90% ethanol, 0.25M NH₄OAc, 100% ethanol, 100% ethanol. The slides were air dried and then used for autoradiography.

Slides were dipped in Ilford K5 gel form emulsion diluted 2:3 with 1.7% (v/v) glycerol, at 43°C, gelled on a glass plate on ice for 15 minutes and then air dried at room temperature for 2 hours. Slides were then dried in a desiccated box at room temperature, for 6-14 hours, and the box placed at 4°C for the exposure period. Slides were exposed for between 5 and 9 days. At 20°C the slides were developed for 2 minutes in Kodak D19 developer, stopped in 1% acetic acid, 1% glycerol in water for 1 minute and fixed in 30% sodium thiosulphate for 5 minutes. The slides were then washed in water, twice for 10 minutes. Sections were stained with 0.02% (w/v) toluidine blue for 1 minute, dehydrated through an alcohol series, cleared in xylene and mounted under a clean glass coverslip using Permount. Sections were examined and photographed, under bright field and dark ground illumination, using an Olympus Vanox-T microscope.

2.6 Sequence Comparisons

Sequences were compared using the software from the University of Wisconsin Genetics Computer Group (UWGCG) (Devreux et al., 1984). The "Gap" programme was used to find the optimal alignment between two predicted amino acid sequences. This programme finds optimal alignment by inserting gaps to optimise the number of alignments and uses the algorithm of Needleman and Wunsch (1970).

2.6.1 Sequences Used for Comparisons

The sequences used for the comparisons in figure 3.3 are from the sources in parenthesis: Hox 2.5 and Hox 2.4 (Graham et al., 1989); Hox 2.6 and Hox 2.7 (Graham et al., 1988); Hox 2.1 (Krumlauf et al., 1987); Hox 2.2 (Hart et al., 1987); Hox 2.3 (Meijlink et al., 1987); Hox 1.7 (Rubin et al., 1987); Hox 1.1,
Hox 1.2, Hox 1.4 and Hox 1.5 (Duboule et al., 1986); Hox 1.3 (Odenwald et al., 1987); Hox 1.6 (Baron et al., 1987); Hox 3.1 and Hox 3.2 (Brier et al., 1988); Hox 6.1 (Sharpe et al., 1988); Hox 4.1 (Lonai et al., 1987); Hox 5.1 (Featherstone et al., 1988); Hox 5.2 and Hox 5.3 (Dolle and Duboule, 1989); Abd-B (Regulski et al., 1985); abd-A (Akam et al., 1988); Ubx and Antp (Scott and Weiner, 1984); Scr (Kuriowa et al., 1985); Dfd (Regulski et al., 1987); zenI (Rushlow et al., 1988); lab /F90-2 (Hoey et al., 1986, Mlodzik et al., 1988).

2.7 Hox 3.1 probe

The Hox 3.1 probe that was used in part of this work was a gift from Dr. Stephen Gaunt and is described in Gaunt (1988).
Chapter Three - The Murine and Drosophila Clusters Are Evolutionary Related

3.1 Introduction

It became apparent as murine homeobox sequences were being cloned that these motifs were clustered close together on the genome (Colberg-Poley et al., 1985, Hart et al., 1985). Currently it seems that the mouse genome contains at least 27 "Antp like" genes. These genes are organised into four clusters, termed Hox 1, Hox 2, Hox 3 and Hox 5 and are located on chromosomes 6, 11, 15 and 2 respectively (Bucan et al., 1986, Hart et al., 1985, Brier et al., 1986, Featherstone et al., 1988). The Hox clusters do not appear to map at the same place as any previously described mouse loci. The clusters were numbered in order of their discovery, thus Hox 1 was the first cluster to be discovered. Although there have been reports of a Hox 4 cluster (Lonai et al., 1987) and a Hox 6 cluster (Sharpe et al., 1988) these clusters have been reassigned to existing clusters. Hox 4 is part of Hox 5 and Hox 6 is part of Hox 3 (Schughart et al., 1989). Each member of the clusters is designated by decimal numbers, such as Hox 2.6. The decimal numbers were initially meant to reflect the linear order of the Hox loci along the chromosome (Martin et al., 1987).

The Hox 2 cluster contains at least 9 members to date, as shown in figure 3.1. This figure also shows the overlapping cosmid clones that span the cluster and that eight members have been cloned within a yeast artificial chromosome (YAC). As with other murine homeobox clusters the Hox 2 cluster is contained within a relatively small piece of DNA, with the 9 genes lying in a 150 kb stretch. All of these genes, and indeed other murine genes, share a common genomic structure and this is illustrated with the Hox 2.6 gene. Typically these genes are small. The Hox 2.6 gene comprises of two exons within 3.2 kb with the homeobox being located in the second exon (Figure 3.1). The finding of the homeobox in the last exon is also typical of the Drosophila homeobox genes, although in some cases it will be split by an intron (e.g. Hoey et al., 1986). The first exon also contains a conserved region, the hexapeptide, that is found in many homeobox genes from both mice and flies.

While some of the Drosophila homeobox genes are enormous, such as Antennapedia which is greater than 100 kb (Schneuwly et al., 1986), the murine genes tend to be relatively small at around 3 kb. The major transcript of the Hox 2.6 gene is 2.4 kb and contains a long 3' untranslated region. This large untranslated region contains an AUUUA motif which may be involved in post-transcriptional control (Shaw and Kamen, 1986). The difference in gene size between Drosophila and mouse is also reflected in the size of the proteins with many of the fly proteins containing runs of repetitive amino acids and
Figure 3.1 - The organisation of the Hox 2 cluster of murine homeobox genes and a diagrammatic representation of the Hox 2.6 gene structure. The overlapping cosmid clones and YAC clone that span much of the cluster are illustrated. The predicted Hox 2.6 coding sequence is shown.

Figure 3.2 - Gap alignment to identify regions of identity between different protein sequences. A-Sequence comparisons between murine Hox genes from the same cluster, Hox 2.1 vs Hox 2.6, and from the Hox 1 and Hox 2 clusters, Hox 2.3 vs Hox 1.1, Hox 2.2 vs Hox 1.2, Hox 2.1 vs Hox 1.3 and Hox 2.6 vs Hox 1.4. B- Sequence comparisons between members of the murine Hox 2 cluster and homeobox genes from the zebrafish, Hox 2.1 vs zf21 and Hox 2.2 vs zf22, and homeobox genes from Xenopus, Hox 2.3 vs XlH 2. In all comparisons the bars represent identities between the two sequences that are being compared.
being considerably larger (e.g. Regulski et al., 1987). Conceptual translation of the Hox 2.6 sequence produces a 250 amino acid protein with an in frame hexapeptide and homeodomain. Like many of the other homeodomain proteins, Hox 2.6 is rich in prolines and in serines and as such has many potential sites for post-translational modification.

3.2 The mouse Hox clusters are related, possibly by duplication

It has been noted that the genes of Hox 1 and Hox 2 share similar intergenic spacing and all are transcribed in the same 5' to 3' direction (Graham et al., 1988). There are many examples of genes in Hox 2 being more related to genes in Hox 1 than to other Hox 2 genes. If one compares the predicted amino acid sequences from two genes from the same cluster, Hox 2.6 and Hox 2.1 for example, the only notable region of identity is at the homeodomain. This is shown in figure 3.2A where amino acid sequences of a number of genes have been compared and identities are shown by bars and the homeodomain is marked by a line. In this comparison it is clear that the greatest density of bars is at the homeodomain. Yet if one compares members of Hox 2 with members of Hox 1 it is clear that the genes fall into Hox 1 - Hox 2 pairs.

Comparisons of the homeodomains allows one to separate generally the genes into pairs with a Hox 1 member and a Hox 2 member (Figure 3.3). One can clearly see that these pairs show high levels of sequence identity with each other, sharing common specific amino acid changes when compared to Antp. The pairs are Hox 2.5/1.7, 2.3/1.1, 2.2/1.2, 2.1/1.3, 2.6/1.4, 2.7/1.5 and 2.9/1.6. Where full amino acid sequence is available it can be demonstrated that those genes identified as being related through homeodomain comparisons also exhibit related regions outside the homeodomain. This is shown for Hox 2.3 and 1.1, for 2.2 and 1.2, for 2.1 and 1.3 and for 2.6 and 1.4 (Figure 3.2A). Clearly there is a greater density of bars at the N-termini and just upstream of the homeodomain at the hexa-peptide. There is also extended identity around the homeodomains themselves. Taken together these observations would support the idea that Hox 1 and Hox 2 are related by a duplication of a cluster of at least 7 homeobox genes as is illustrated in figure 3.4. No mouse Hox 1 counterparts have yet been found for Hox 2.4 or 2.8. Further characterisation will reveal the existence, or not, of these and other genes and help define the duplication, and the extent of divergence between the two clusters. Recent information would seem to suggest that while a large portion of Hox 1 and Hox 2 are related through duplication the whole clusters do not appear to be related in such a manner. It appears that the Hox 1 cluster is more extensive than Hox 2 and possesses more 5' members than Hox 2 (P.Gruss
Figure 3.3 - Amino acid sequence comparisons of the mouse and Drosophila homeodomains showing the existence of related subgroups. The Antp sequence at the top was used as a basis for all comparisons. Dashes represent identical amino acids. The separation into subgroups is based on the changes shared by the various members. References for the sequences used in these comparisons are detailed in material and methods.

Figure 3.4 - A schematic alignment of the Hox 1 and Hox 2 clusters. Hox 1 is located on chromosome 6 and Hox 2 on chromosome 11.
Figure 3.5 - Comparison between the related murine homeodomain protein sequences, Hox 2.6 and Hox 5.1. The Hox 2.6 sequence from Graham et al., 1988 and the Hox 5.1 sequence from Featherstone et al., 1988.

Figure 3.6 - Generalised structure showing characteristic features and conserved regions of a typical vertebrate homeodomain protein. This diagram is based on several vertebrate sequences. (AA) amino acid.
The relationship of the other two clusters to the Hox 1 and Hox 2 clusters is more obscure but again one can find related members at relatively similar positions in each cluster. In figure 3.3 one can note the relationship between the Hox 2 homeodomains and their closest murine counterparts. This comparison clearly points out a lot of significant similarities between many of the mouse homeodomains, and allows one to separate many of the mouse genes into subfamilies of related sequences. These subfamilies are defined by common amino acid changes and the position of the changes (when compared to Antp) that are shared within a family. On this basis Hox 2.5, 1.7, 3.2 and 5.2 all form a subfamily and the mouse homeobox genes fall into eight such subfamilies (Figure 3.3).

The relatedness of genes within a subfamily is strengthened by the fact that members of a subfamily have a high degree of identity in multiple regions over their entire predicted protein sequences. Full sequences are available for Hox 1.4, Hox 2.6 and Hox 5.1 and comparison shows that the related areas are the N-termini, the hexapeptide and the homeo domain and sequences flanking. This is shown for Hox 2.6 and Hox 1.4 in figure 3.2 A and for Hox 2.6 and Hox 5.1 in figure 3.5. The Hox 5.3 and Hox 2.8 genes have not been placed within a subfamily since it is possible that these two genes represent members of two other subfamilies and are not members of the other eight subfamilies. This question will be clarified through more cloning and sequence work. Still it would appear that all of the murine homeobox gene clusters are related to each other.

Comparisons between related genes, in the same subfamily, always shows the N-terminus, the hexapeptide and the homeodomain to be the related regions. The conservation of these regions may be due to functional constraints. It is known that the homeodomain is a DNA binding region but roles have not been suggested for the N-terminus and the hexapeptide. For murine homeodomain proteins the middle unique portion is often rich in prolines and serines. Thus a typical homeodomain protein can be considered as shown in figure 3.6.

3.3 Similarities between mouse Hox complexes and those of other vertebrates

It has been shown that the human homeobox genes are organised as in the mouse (Boncinelli et al., 1988) and that these genes show very high degrees of sequence identity to their mouse homologues. For example the murine and human Hox 5.1 genes show 93% amino acid identity over the whole coding region (Featherstone et al., 1988). As yet there has been no gene that has been described in one organism, either mouse or human, but not the other.

More distantly related vertebrates would also seem to organise their
homeobox genes in a similar fashion. One can note from figure 3.2B that the murine Hox 2.3 sequence displays extensive identity over the whole protein sequence with a *Xenopus* X1H2 sequence (Fritz et al., 1989). This may suggest that the Hox 2 genes are found in many vertebrate species. More convincing evidence for this suggestion comes from an analysis of sequences from the zebrafish, *Brachydanio rerio*. This animal has been found to possess two closely linked genes that would appear to be homologues of Hox 2.1 and Hox 2.2 (Njolstad et al., 1989). A comparison between the murine Hox 2.1 and the zebrafish zf-21 and between the Hox 2.2 gene and the zebrafish, zf-22 is illustrated in figure 3.2B. The Hox 2.1 pair of genes share 81% sequence identity and the Hox 2.2 pair share 70%, at the amino acid level (Njolstad et al., 1988, Njolstad et al., 1989). These zebrafish genes are closer to these mouse genes than to any other murine genes so far described, suggesting that they are directly related. The genome of *Brachydanio rerio* also contains two linked genes that are most closely related to the murine Hox 6.1 and 6.2 genes (Njolstad et al., 1989). It would seem that the zebrafish may possess true homologues of both the Hox 2 and Hox 3 clusters of homeobox genes.

To be sure of this point, and to determine the existence, or not, of zebrafish Hox 1 and Hox 5 loci, more characterisation of zebrafish homeobox genes is necessary. Since the evolutionary distance between teleosts and mice is relatively large, one would be tempted to suggest that the common features that exist between the fish and in the mouse will hold true for all vertebrates.

### 3.4 The mouse and Drosophila homeobox gene complexes have common features of organisation

Many of the *Drosophila* homeotic genes have been found to be clustered in two separate but adjacent complexes, Antennapedia (ANT-C) and Bithorax (BX-C) (Lewis, 1978, Kaufman et al., 1980, Sanchez-Herrero et al., 1985). Several homeotic lethal complementation groups have been identified in the ANT-C. These include *Antennapedia* (*Antp*), *Sex combs reduced* (*Scr*), *Deformed* (*Dfd*), *proboscepedia* (*pb*) and *labial* (*lab*). The BX-C has three essential domains of homeotic function which are *Abdominal-B* (*Abd-B*), *abdominal-A* (*abd-A*) and *Ultrabithorax* (*Ubx*). If one includes the *Drosophila* homeotic genes of the BX-C and ANT-C in the homeodomain comparisons one can note that many of these sequences show similarity to one subfamily. The comparisons between *Abd-B* and Hox 2.5, *Scr* and Hox 2.1, *Dfd* and Hox 2.6, and *lab* and Hox 2.9 show the highest degrees of sequence identity. The *Drosophila* genes *abd-A*, *Ubx* and *Antp* are all so similar in sequence that it is difficult to assign these genes to any of the other murine subfamilies, which are represented by Hox 2.4, 2.3 and
Figure 3.7 - Diagrammatic representation of the relationship between Hox 2.6 and the Drosophila Dfd. The shaded areas represent amino acid identity, and dashed lines demark related areas in the two proteins.

Conservation Between Homeobox Complexes

Figure 3.8 - Representation of the relationship between Drosophila and murine homeobox gene clusters. The sequence comparisons from figure 3.3 were used to align the clusters into an evolutionary related network. The different members of a related subfamily are represented by a vertical row of boxes. Filled boxes in the mouse clusters represent sequenced homeodomains, open boxes indicate identified but not as yet sequenced mouse homeodomains, and hatched boxes indicate predicted genes, based on analogy to other clusters. The numbers in parenthesis represent the chromosomal location of these clusters. The numbers above each box were those that were originally assigned as the genes were isolated. Hox 6.1 and Hox 6.2 and Hox 4.1 have been reassigned to existing clusters on chromosomes 15 and 2 respectively (Schughart et al., 1989). Their original nomenclature has been kept until they were renamed.
2.2. The *Drosophila* gene *pb* would seem to fit into both the murine subfamilies represented by Hox 2.7 and by Hox 2.8 (Cribbs and Kaufman pers. comm.).

Hox 2.6 displays several regions of identity, outside of the homeodomain, when compared with the *Dfd* protein (Figure 3.7) (Graham et al., 1988) and the *Scr* protein (leMotte et al., 1989) shares identity with Hox 2.1, particularly in the amino-terminal region, where 10 of the first 13 amino acids are the same. Mlodzik et al., (1988) have found extended regions of relatedness between *lab* and Hox 1.6 (LaRosa and Gudas, 1988). Therefore, where protein sequence is available, mouse and *Drosophila* genes identified as subfamily members, based on similarities within the homeodomain, also display some identity in other domains. This is obviously also true for other subfamily members and, for example, Hox 5.1 displays the same type of relationship to *Dfd* as Hox 2.6 (Featherstone et al., 1988). Again the conserved regions are at the N-terminus and the hexapeptide which would reinforce suggestions that these regions are conserved because they are functionally important.

3.5 A common ancestor for mouse and *Drosophila* homeobox gene complexes

One striking feature of these comparisons is that if one aligns the mouse and *Drosophila* subfamilies, then the physical order of the *Drosophila* homeobox genes along the chromosome and that of their closest murine genes is identical (Figure 3.8). These results suggest that the murine and *Drosophila* homeobox gene clusters are derived from a common ancestor. Since mice and *Drosophila* are representatives of the deuterostomes and protostomes respectively it is suggested that this ancestral cluster arose before the divergence of these lineages. In the above comparisons the best overall identities between mouse and *Drosophila* are in those genes (*lab*, *pb*, *Dfd*, *Antp*, and *Abd-B*) predicted to be the most ancient by Akam et al., (1988). In addition, the identity between *Scr* and Hox 2.1 argues that *Scr* was part of the ancestral cluster. Therefore it is suggested that there were at least six homeobox genes in the ancestral cluster. Figure 3.9 shows one possible model from which one could derive the *Drosophila* and the murine Hox 2 clusters from the putative ancestral cluster.

Akam et al., (1988) have recently proposed that the earliest ancestor for the myriapod-insect lineage possessed a homeobox cluster with distinct genes related to *lab*, *pb*, *Dfd*, *Antp* and *Abd-B*. Subsequently by duplication and divergence the *Antp* gene gave rise to the *Scr* gene and then, more recently, to the *abd-A* and *Ubx* genes. It is interesting that the genes defined by these workers as being the most ancient are also thought to be ancient from the comparisons between *Drosophila* and mouse. Yet the comparisons that are described here would suggest that the cluster postulated by Akam et al., (1988) in fact existed long before any ancestor for the myriapod-insect
Figure 3.9 - A possible scheme for the evolution of the *Drosophila* and murine clusters from a common ancestral cluster. This diagram is based on sequence information that is currently available.
lineage. Also considering the relationship between Scr and Hox 2.1, outlined above, it is postulated that this gene was part of the ancestral cluster. In both models the specific abd-A, Ubx and Antp genes are thought to have been generated more recently. It is noteworthy that it is difficult to assign any of these three genes to the murine Hox 2.4, 2.3 and 2.2 genes. This could be a result of independent duplication, in each of the major lineages, of an ancestral "Antp" like gene. So the fact that the three Drosophila genes and the three murine genes show a high degree of sequence identity reflects their separate derivation from a common ancestral gene. In the case of pb, which shows a high degree of identity with Hox 2.7 and 2.8 (Cribbs and Kaufman pers. comm.), it is thought that there was a duplication of the ancestral pb gene in the deutostome lineage to generate these two Hox 2 genes. Such independent duplications has been postulated for the engrailed genes (Doleki and Humphreys, 1988, Joyner and Martin, 1987).

It is also important to note that the murine genome contains, not one, but four homeobox gene clusters which are all related to each other. Drosophila has a single homeobox cluster and if it is assumed that there has been no secondary loss of homeobox gene clusters from arthropods it would seem that the four murine clusters are directly or indirectly evolved from the single ancient ancestral cluster. In each of the four clusters, the relative position of each subfamily member and the physical order and the spacing of the genes are very similar. The alignment, shown in figure 3.8, also highlights the fact that not all genes or subfamilies are represented in all the murine clusters. This may suggest that either some clusters arose at different stages during the evolution of the clusters or resulted from partial duplication or there has been sufficient divergence to make some genes undetectable. Interestingly the genes that are absent from Hox 3 and Hox 5 are among those that are thought to be derived from Antp and therefore evolutionary more recent.

The sequence identity between the mouse and the Drosophila genes should not be thought of as representing the existence of true vertebrate homologues of Drosophila homeotic genes. Some subfamilies have four murine genes related to a given Drosophila gene. As such it would be hard to establish a precise murine equivalent.

These comparisons suggest that the Drosophila and murine homeobox clusters arose from a common ancestral cluster that existed prior to the split between the protostomes and the deutrostomes. This is supported by identity between the mouse and Drosophila proteins not only in the homeodomain but also, where information is available, in other regions outside the homeo
domain (Figure 3.7, Graham et al., 1988). One difficulty with this argument is that the Drosophila genes are divided in two separate clusters although both are on chromosome three. However, in a more primitive insect, the red flour beetle, Tribolium castaneum, the ANT-C and BX-C genetic equivalents are found in a single homeotic cluster, HOM-C (Beeman et al., 1989). Therefore, ANT-C and BX-C may originally have been one cluster that has been separated by more recent evolutionary events occurring in some insects.
4.1 General aspects of murine homeobox gene expression

Murine homeobox genes are not expressed in all tissues at all times in embryogenesis but rather show restricted patterns of expression. The fact that they exhibit spatially and temporally restricted patterns make them all the more interesting as this would be one of the characteristics that one would expect to find in genes that were involved in controlling important developmental processes, such as regional specification or various inductive interactions. That is not to say that all genes that are important in controlling such processes will be restricted in their patterns of expression. Yet if a gene were to show a general expression pattern it would also be likely that this gene would have a house-keeping function.

It is useful to consider the patterns of expression of one of these genes at a frozen time point. The time point that is often chosen is at 12.5 days of gestation since this represents the time at which expression levels peak for many of these genes. It is also a convenient stage for the analysis of the expression patterns in a number of the organ systems that are being established. All of these genes have been described as being expressed in ectodermal and mesodermal derivatives but not in endodermal derivatives. Their expression is also restricted within mesodermal and ectodermal tissues. This general pattern is illustrated in figure 4.1 with the Hox 2.2 gene.

Within ectodermal derivatives it is apparent from figure 4.1 that Hox 2.2 is detected in components of both the central and peripheral nervous systems. Expression is found from most posterior regions of the spinal cord and extends anteriorly all the way along the spinal cord until expression stops in the hindbrain. While there is no posterior limit of expression within the central nervous system, there is a very clear anterior limit - indicated with an A. This anterior limit is interesting because at 12.5 days of gestation it does not correspond to any obvious morphological boundary. The gene is not expressed in the midbrain or forebrain. The section in figure 4.1 A demonstrates the expression of this gene in the dorsal root ganglia. These are peripheral nervous system components, paired ganglia that are arranged in a repeating pattern lateral to the spinal cord and are derived from the neural crest. Hox 2.2 can be seen to be detected in the most posterior ganglia up to the most anterior ganglia.

Hox 2.2 can be found to be expressed in many mesodermal structures. Expression in prevertebrae extends from most posterior regions to the 7th
Figure 4.1 - *In situ* hybridisation of the Hox 2.2 gene in the 12.5 day mouse embryo. Both bright field (left) and dark ground (right) are shown. A number of structures are labelled: HB- hind brain, SC- spinal cord, PV- prevertebra, H- heart, L- lung, Li- liver, S- stomach, K- kidney, DRG- dorsal root ganglia. The anterior limit of expression in the central nervous system is marked (A) as is that in the prevertebrae (>).
Figure 4.2 - Dark ground photograph of an in situ hybridisation to a 12.5 day embryo sagittal section with a sense probe. One can note that the liver (li) and the blood cells (*) appear to be positive. There would also appear to be some signal around the edge of the section.

Figure 4.3 - Schematic representation of the fate map of a Drosophila blastoderm stage embryo. Anterior is to the left and dorsal to the top. The black bars represent the domains of expression of the homeotic genes. The segments are indicated at the top (Md, mandibulary; Mx,maxillary; Lb, labial segment; T1-T3, thoracic segments; A1-A10, abdominal segments). the primordia of the central nervous system is also indicated (CNS). Dfd, Deformed; Scr, Sex combs reduced; Antp, Antennapedia; Ubx, Ultrabithorax; iab-2 infraabdominal-2; iab-7, infraabdominal 7; and cad, caudal
prevertebrae - marked with an arrow in figure 4.1 B. This is analogous to the situation that is observed in the CNS, where one observes expression from posterior regions up to an anterior limit. One also detects transcripts in the viscera. It is important to note that in organs of composite origin, such as the lung, this gene, and others, are only expressed in the mesodermal components. In figure 4.1 A one can observe the expression of Hox 2.2 in the mesenchyme of the lung, a mesodermal component, and not in the inner epithelium which has an endodermal origin. Expression of Hox 2.2 can also be seen in the mesenchyme of the stomach which although weak has been confirmed by northern analysis (Figure 4.1 A, N.Papalopulu pers. comm.). Expression is not restricted to mesenchyme and in the metanephric kidney Hox 2.2 transcripts are expressed abundantly in the epithelial cells (Figure 4.1 A, Figure 4.10). Noticeably there are other mesodermal organs which do not express the gene. This is demonstrated in figure 4.1 with the heart and expression of Hox 2.2 has not been detected in the developing limbs (see later).

In figure 4.1 one can note that the liver appears to be highlighted with the Hox 2.2 probe. This in fact reflects an in situ hybridisation artefact. It is clear from figure 4.2, where a similar section has been probed with a sense probe, which should not detect any mRNA, that the liver is also highlighted. In figure 4.2 it is clear that blood cells also appear as if they were positive even when one uses a sense probe. Since the liver and the blood cells are both highlighted even when one uses a probe that does not detect mRNA one can rule out these signals as being due to artefacts related to this technique. In the case of the liver it has been shown, by northern blotting that there is no expression of Hox 2.2 (N.Papalopulu pers. comm.). The 12.5 day mouse embryo liver is a major site of haemopoiesis and it would seem that the apparent positive signals related to the blood cells and the liver is due to the fact that blood cells non-specifically stick the riboprobes. In cases where expression in an organ such as the liver is not clear from in situ hybridisation the results are backed up by the more sensitive method of northern blot analysis. All of the northern blot analysis was performed by Nancy Papalopulu. One should also note that areas around the edge of the signal also appear to be positive but again since these regions are also highlighted with the sense probe it would seem that this is due to the non-specific adherence of the riboprobes.

4.2 Relationship between the clustering of homeoboxes and their expression

The clustering of homeobox genes is a general phenomenon having been found in every case so far analysed from mouse to zebrafish to Drosophila. In
Drosophila one important feature of the clusters is that the physical order of the homeotic genes is identical to the order in which these genes are expressed along the anteroposterior axis of the embryo during development (Harding et al., 1985, Akam, 1987). The domains of expression along the axis also correspond directly to the segments that are affected by mutations in these genes (Harding et al., 1985). The expression of the homeotic genes in the blastoderm embryo is shown diagrammatically in figure 4.3. Note that Abd-B is infraabdominal 7 and abd-A is infraabdominal 2. The telomeric most gene of BX-C, Abd-B, is expressed caudally while the genes of the ANT-C which lie towards the centromere, such as Antp, Scr and Dfd, are more rostrally expressed. This pattern is maintained with each more centromeric gene exhibiting a more rostral domain of expression i.e. Dfd is expressed more rostrally than Scr. While many of the homeotic genes have patterns of expression that are linked to metamerism this is not true of all the homeotic genes. For example labial is expressed in non-segmented regions of the Drosophila embryo (Hoey et al., 1986). These studies and others have lead to the suggestion that the homeotic genes are responsible for the specification of position along the anteroposterior axis (Slack, 1984). A given parasegment, or group of cells, would perceive its position along this axis through expression of homeotic genes.

As has been discussed there is a group of genes which are homologous to the Drosophila homeotic genes, as defined by genetic means, in the red flour beetle, Tribolium castaneum, HOM-C (Beeman et al., 1989). In this insect it is also clear that there is also a correlation between the physical order of the genes and the order, along the anteroposterior axis, of the segments affected by mutations in these genes. Thus in these two organisms, at least, the clustering of homeotic genes would appear to be important and to reflect aspects of there function. Consideration of the mutant phenotypes would suggest that these genes are functioning in the specification of anteroposterior position.

In the previous chapter it was suggested that the murine and Drosophila homeobox clusters are both derived from a common ancestral cluster. It is important to determine if this common evolutionary origin is also reflected in the expression patterns of the members of a murine homeobox cluster. If the mouse homeobox gene cluster also displays a correlation between the physical order of the genes and their expression along the rostrocaudal axis then one would expect such a correlation to be universal within the protostomes and the deutrostomes. It is also interesting that the murine Hox 2 cluster is so small and has nine genes within only 150 kb of DNA and that this stretch of the
mammalian genome is devoid of repetitive elements. Therefore in situ hybridisation has been used to determine the patterns of expression of the members of the Hox 2 cluster in embryogenesis. The 12.5 day mouse embryo has been extensively used because this represents the stage at which the members of Hox 2 show the highest levels of expression.

4.3 Hox 2 probes used in in situ hybridisation

Portions of seven of the members of Hox 2 were used in the following in situ hybridisation experiments. All probes were subcloned into riboprobe vectors, either the pGEM 1 vector (Promega Biotech) or into the Bluescript vector (Stratagene). The multiple cloning sites of these vectors are flanked by prokaryotic promoters that allow in vitro transcription using the appropriate RNA polymerase, T7 or T3. The Hox 2.1 probe that was used is an 800 bp SacI-EcoRI fragment whose 5’ SacI end is at the N-terminal coding region and whose 3’ EcoRI sites lies in the 5’ half of the homeo domain. This fragment was subcloned into the pGEM 1 vector (Krumlauf et al., 1987). All other fragments for each of the genes were subcloned into the Bluescript vector. In all cases the DNA fragment of interest that was to be subcloned was isolated from an agarose gel after electrophoresis by electrophellation. These isolated fragments were ligated into vectors that had been digested with the appropriate restriction enzyme. The ligation products were transformed into JM 109 cells and plated on L-agar containing the appropriate selectable marker. The pKS vector contains the lacZ a-peptide coding region in its poly-linker and as such after transformation into the JM 109 cells which carry the M15 mutation in the lacZ gene on can screen for recombinant plasmids by screening for beta-galactosidase activity. The screening for the beta-galactosidase activity involves use of the chromogenic substrate X-gal. Recombinant colonies appear white while non-recombinant colonies are blue. Positive clones were identified either by the above blue-white selection followed by restriction digests of samples of colonies after the plasmid DNA was mini-prepped or by colony hybridisation. The genomic structure of these genes has been determined by people in this laboratory and by many other workers in this field.

The Hox 2.5 probe was a 400 bp BgIII-DraI fragment that spanned the homeobox. This fragment was derived from a 1kb SacI subclone from the cosmid pCos5.4 (see Figure 3.1). The Hox 2.4 subclone that was used to generate the Hox 2.4 probe was produced by N.Eager in the laboratory and consisted of a 1kb SacI genomic fragment that spanned the homeobox and the regions 3’. A 5kb Hox 2.3 fragment was isolated from the cosmid pCos5.4 (Figure 3.1) and from this a smaller region was subcloned to be used as a probe. This was a 800 bp EcoRI-
BamHI fragment that started in the intron just upstream of the second exon, EcoRI site, and terminated in the 3′ untranslated region (See Meijlink et al., 1987, for full map). The Hox 2.2 probe was a 1.2 kb EcoRI-SacI fragment that spanned the homeobox and was subcloned directly from pCosH15 (Figure 3.1). This fragment was identified by southern blotting and probing with a 220 bp EcoRI-XhoI fragment containing the Antp homeobox (McGinnis et al., 1984). The EcoRI site lies in the intron and the SacI site in the 3′ untranslated region as is shown in the map of Schughart et al. (1988). The Hox 2.6 probe was derived from a cDNA clone, cDNA 7, and is an 800 bp fragment that starts at a SalI site close to the beginning of the long open reading frame and extends to a BglII site in the homeobox. The Hox 2.7 probe originated from a genomic clone and has a 5′ BglII site in the homeobox and a 3′ site in the untranslated region.

4.4 Ordered expression of Hox 2 genes in the central nervous system (CNS)

Using all the Hox 2 probes, overlapping but distinct domains of expression have been observed for each gene in the CNS. In view of the absence of a clear caudal boundary of expression, this work has focused on examining the different rostral boundaries of expression for each member of the Hox 2 complex. The same rostral boundary of expression was consistently observed in all 12.5 day mouse embryos. All seven genes have been directly compared on serial sections, using relative distances from morphological boundaries to map the expression boundaries. Starting with Hox 2.5 one finds that as one progresses through the Hox 2 complex in a 5′ to 3′ direction, the limit of expression of each gene in the CNS becomes progressively more rostral. Figure 4.4 shows that for the 5′ most gene, Hox 2.5, the limit of expression maps within the spinal cord at the level of the third/fourth cervical prevertebra. The boundary of the adjacent Hox 2.4 gene is within the posterior myelencephalon, as are those of the more 3′ genes Hox 2.3 (not shown) and Hox 2.2 (Figure 4.4). The difference in the rostral boundary between Hox 2.5 and Hox 2.4 is large by comparison with the difference between Hox 2.2 and Hox 2.4. However each gene does have a successively more rostral limit (Figure 4.4).

Figure 4.5 illustrates, in serial sections of an embryo, that this general trend continues through the complex. Domains of expression for members 3′ of Hox 2.2 extend more rostrally within the myelencephalon, and there is a relatively large difference between the rostral boundaries of Hox 2.2 and Hox 2.1. The successively more rostral expression in the hindbrain for Hox 2.1-2.7 is shown in figures 4.5 and 4.6. A summary of these in situ results is
Figure 4.4 - Comparison of the patterns of expression for Hox 2.5, Hox 2.4 and Hox 2.2 in the 12.5 day mouse embryo. The top row displays *in situ* hybridisation on near adjacent sections showing bright fields of the entire embryo. Below are higher powered views of these sections focusing on the anterior limits of expression. These are represented in both bright field and dark ground pictures.
Figure 4.5 - Anterior boundaries of expression in the central nervous system for six members of the Hox 2 cluster as revealed by in situ hybridisation. The areas (A and B) of near adjacent sections photographed in high power are illustrated in the bright field picture of the whole 12.5 day embryo at the bottom. Area A was used for Hox 2.7, Hox 2.6 and Hox 2.1 while the more posterior area B was used for Hox 2.2, Hox 2.4 and hox 2.5. Dark ground photographs are shown for each gene as indicated at the top and bottom of the pictures. Bright field pictures of A and B are shown at the left to aid identification of boundaries.
Figure 4.6 - Patterns of expression of the Hox 2 genes that show the most anterior limits of expression in the central nervous system. A bright field photograph of a 12.5 day mouse embryo is shown at the top to indicate the area pictured in higher power. The dark ground photographs show the anterior limits of expression in the hind brain for Hox 2.7, Hox 2.6 and hox 2.1. Probes for each section are indicated at the right.

Figure 4.7 - Diagrammatic representation of the anterior boundaries of expression in the central nervous system for the Hox 2 cluster, and their correlation with gene position in the cluster. A bright field photograph of a 12.5 day mouse embryo is shown below the cluster. Arrows indicate the anterior boundaries in expression detected in near adjacent sections from this embryo. The limits were established using three or four sections for each gene and positioned by comparison of relative distances from morphological structures. P- posterior, A- anterior.
illustrated in figure 4.7. These findings would suggest that the position of a gene within the Hox 2 cluster reflects its relative domain of expression along the rostrocaudal axis of the animal in the CNS.

4.5 Correlation between gene order and expression is also evident in other ectodermal derivatives

Other ectodermal derivatives, such as neural crest and ectodermal placode structures, display expression of members of the cluster. The results presented in this section are from 12.5 day old mouse embryos. The expression of Hox 2 genes is particularly noticeable in the peripheral nervous system. The dorsal root (spinal) ganglia, which are neural crest derived, exhibit expression of every member of Hox 2, from Hox 2.5 to 2.7, for example see figure 4.1 A for Hox 2.2.

As has been described for the central nervous system, more anterior components of the peripheral nervous system, such as the vagal (Xth)/glossopharyngeal (IXth) complex, also show differential expression of members of Hox 2. The components of this complex are derived from the neural crest and from ectodermal placodes (Verwoerd and van Oostrom, 1979, D’Amico-Martel and Noden, 1983). Figure 4.8 shows that the 5’ most member of the Hox 2 cluster, Hox 2.5, is not expressed in these cranial nerves while the more 3’ members are. One can note the expression of Hox 2.2 and 2.1 in the inferior ganglion of the vagal (Xth) nerve, the nodose ganglion, but not in the petrosal ganglion, which is the inferior ganglion of the glossopharyngeal (IXth) nerve. This is demonstrated with the Hox 2.1 gene in figure 4.8 B. Clearly there is hybridisation to the nodose but not the petrosal ganglion. An analysis of the position of these two ganglia in adjacent hybridised sections may also illustrate the expression of Hox 2.7 in the more rostral inferior ganglion of the glossopharyngeal nerve, the petrosal ganglion. Expression of any of these four genes cannot be detected in superior ganglia of either the vagus or glossopharyngeal nerve are formed from more rostral regions than the inferior ganglia (Figure 4.8 A and for Hox 2.7 data not shown) (Altman and Bayer, 1982). The trigeminal (Vth) nerve which lies more rostrally still than these nerves does not express any member of Hox 2. The rostral boundary of Hox 2.7 expression in the hind brain region is also apparent in figure 4.8 A.

Cranial neural crest not only contributes to nervous tissue but is also involved in producing skeletal and connective tissue (Le Douarin, 1982). Figure 4.9 A shows the pharyngeal region that has been probed with Hox 2.5 and Hox 2.7 on adjacent sections. The mesenchyme of the thyroid has a neural
Figure 4.8 - Differential expression of members of Hox 2 in the vagal/glossopharyngeal nerve complex. A- Adjacent sections probed with four members of the Hox 2 cluster, Hox 2.5, Hox 2.2, Hox 2.1 and Hox 2.7 are shown in dark ground. Two Bright field views of this region are shown to allow identification of structures. The probe used for each section is indicated at the top. B- High powered photograph showing Hox 2.1 expression in the nodose but not in the petrosal ganglion. X- Xth (nodose) ganglion, IX- IXth (petrosal) ganglion, O- otocyst, S- superior ganglia of the vagal/glossopharyngeal complex, pg- petrosal ganglion. ng- nodose ganglion
Figure 4.9 - Expression of Hox 2 genes in some neural crest derivatives. A- *In situ* hybridisation of Hox 2.5 and Hox 2.7 to adjacent sections reveals the expression of Hox 2.7 but not of Hox 2.5 in the thyroid gland. The expression patterns of these two genes are shown in dark ground and the bright field photograph at the left allows orientation. The probe used for each section is marked above. B- Hox 2.1 expression in the enteric ganglia. G-gut, L-lung, Li-liver.
Figure 4.10 - Anterior boundaries of expression of three members of Hox 2 in the prevertebrae. The probes used are indicated above each section. One can note the patterns of expression of Hox 2.2, Hox 2.1 and Hox 2.7 in the dark ground pictures. The bright field photograph shows the anatomical detail of this region of the 12.5 day embryo.
crest origin and it is apparent that these cells are highlighted with the Hox 2.7 probe but not with the Hox 2.5 probe (Le Douarin, 1982). One can also note the more anterior expression of Hox 2.7 in the central nervous system and in the prevertebrae. Figure 4.9 B displays the expression of the Hox 2.1 gene in the enteric ganglia, which are neural crest derived. Hox 2.1 hybridisation pattern in the gut is punctate which would suggest that this probe is picking out the myenteric plexus of the gut. Thus in more anterior neural crest or ectodermal placode derivatives it is the more 3’ members of the clusters that are expressed.

4.6 Ordered differential expression of Hox 2 genes in mesodermal derivatives

Hox 2 genes are also found to be expressed in a range of mesodermal derivatives and one can note differential expression of members of Hox 2 in the same mesodermal cells. This is apparent in the prevertebrae of the 12.5 day mouse embryo. In figure 4.10 one can note the anterior boundaries of the expression of three members of Hox 2 in the prevertebrae. Hox 2.2 is expressed from the most posterior prevertebrae to the eighth prevertebra, Hox 2.1, the adjacent 3’ gene, to the seventh prevertebra and Hox 2.7, the most 3’ gene analysed, is expressed from posterior regions to the first prevertebra (Figure 4.10). Thus again, as in ectodermal derivatives, there is a correlation between the order of the genes and their expression along the rostrocaudal axis.

This sort of differential expression is also apparent in visceral mesoderm. In the lung, which is a relatively anteriorly derived organ, one can observe expression of the more 3’ genes but not of the 5’ genes. Figure 4.11 shows adjacent 12.5 day parasagittal sections probed with Hox 2.5, 2.2, 2.1 and 2.7. It is clear that while Hox 2.5 is not expressed in lung mesenchyme the more 3’ genes, Hox 2.2, 2.1, and 2.7 are. Again one can see a clear restriction of expression in the lung to the mesodermal components and no expression in the endodermally derived epithelium. In figure 4.11 one can also see apparent signal over the liver yet this is due to in situ artefact and these genes have not been found to be expressed in liver by Northern blot analysis (N. Papalopulu pers. comm.).

This mesoderm versus endoderm restriction can also be observed in the stomach and the intestine. Figure 4.12 demonstrates in a 14.5 days transverse section that transcripts for Hox 2.5 cannot be detected in either the gut or the stomach yet those for Hox 2.1, 2.6 and 2.7 can be. In figure 4.12 the expression of Hox 2.6 is not that strong but this result has been confirmed in other embryos (see Graham et al., 1988) and by northern blot analysis. It has
Figure 4.11 - Differential expression of members of Hox 2 in the lung of the 12.5 day mouse embryo. A bright field photograph of an entire embryo is shown at the bottom to indicate the positions of the various anatomical structures. The dark ground pictures illustrating the expression patterns of each gene are displayed and are labelled with the probe that was used. PV- prevertebrae, H- heart, L- lung, Li- liver, S- stomach, DRG- dorsal root ganglia, K- kidney.

Figure 4.12 - Expression patterns of members of the Hox 2 cluster in the stomach and gut of a 14.5 day mouse embryo. The bright field photograph shows a transverse section through a 14.5 day mouse embryo that contains various visceral organs which are labelled. The dark ground photographs demonstrate the expression of the more 3' members of the cluster in the stomach and in the intestine. Each photograph is labelled with the probe that was used. Li- liver, S-stomach, E- epithelium, G- gut
Figure 4.13 - Mesonephric tubules of the 10.5 day embryo displaying expression of the Hox 2.4, Hox 2.2 and Hox 2.6 genes. The dark ground photographs clearly show strong expression of these genes in the mesonephric tubules and the probe used on each section is indicated. The bright field photograph reveals the anatomical structures. S-somites, M- mesonephric tubules.
Figure 4.14 - Expression of Hox 2 genes in the metanephric kidney, developing gonads and in the adrenal gland of a 14.5 day embryo. The expression of the genes in these organs is shown in the dark ground photographs and each gene probe is indicated. The bright field photograph of the transverse section of this 14.5 day embryo displays the positions of organs of interest. AG- adrenal gland, M- degenerating mesonephros, K- metanephric kidney, G- developing gonad.
also been shown that Hox 2.2 can be detected in the stomach (Figure 4.1 A). In both the stomach and the gut these transcripts are not detected in the inner endodermal layers but in the outer mesodermal layers. The expression of Hox 2.1 in the gut is generally more punctate and as has been discussed above this is thought to represent expression primarily in the enteric ganglia which are derived from neural crest (see Figure 4.9 above). Again in figure 4.12 the liver is also highlighted.

Figures 4.13 and 4.14 show expression of some members of the cluster in the mesonephric and metanephric kidneys. Expression of Hox 2.4, 2.2 and 2.6 are shown in the mesonephric tubules, at the hind limb level, at 10.5 days in figure 4.13. There is very strong hybridisation of these probes to the mesonephric tubules and also lower hybridisation to the other regions that are evident in this section. These other regions are the neural tube and the somites which are in the process of breaking down to yield sclerotome and dermamyotome. A more complete analysis of the expression of Hox 2 members in the metanephric kidney is shown at 14.5 days in figure 4.14. In this figure one can note the expression of Hox 2.5, 2.4, 2.2, 2.1, 2.6 and 2.7 in the metanephric kidney. The sections probed with Hox 2.1, 2.6 and 2.7 also reveal the expression of these genes in the adrenal gland, which has a neural crest component. Further analysis is required to determine if all of the Hox 2 genes are expressed in the adrenal gland. Thus all members of Hox 2 are expressed in the kidney.

In figure 4.14 the developing gonad is apparent and one can observe the differential expression of Hox 2 genes. Hox 2.5, 2.4 and 2.2 are not expressed while the remaining 2.1, 2.6 and 2.7 show expression in the gonad. Expression of Hox 2 genes has not been detected in the heart or in the limbs, either fore or hind. Again the liver artefact is seen in some sections.

The expression patterns that have been described for the murine Hox 2 genes are analogous to those of the Drosophila BX-C and ANT-C homeobox gene clusters. In both cases it is evident that the physical order of the genes along the chromosome is related to the expression of the homeotic genes along the anteroposterior axis. Besides being true for the blastoderm stage this relationship is also observed in the embryonic CNS (Harding et al., 1985, Martinez-Arias et al., 1987, Mlodzik et al., 1988). The telomeric most member of the BX-C Abd-B shows expression to the sixth abdominal ganglion, while the adjacent gene, abd-A, in the centromeric direction, expresses to the first abdominal ganglion and the next centromeric gene Ubx reaches the metathoracic ganglion. ANT-C lies towards the centromere and genes within this complex
maintain this trend. *Antp* expresses in the prothoracic ganglion, *Scr* to the second subesophageal ganglion and *Dfd* to the first subesophageal ganglion. The centromeric most member of ANT-C lab has been reported to be expressed more anteriorly than *Dfd* (Mlodzik et al., 1988).

4.7 Discussion: Position of genes in a cluster reflect domains of expression along the rostral-caudal axis for both mouse and *Drosophila*

4.7.1 Ordered expression in ectoderm

It has been shown that all members of Hox 2 are expressed in slightly different, but overlapping, domains in the CNS with respect to the anteroposterior axis (Figures 4.4, 4.5, 4.6). Posterior limits of expression have not been detected for any of these genes in the nervous system. However, as one moves from the 5' most member of the cluster (Hox 2.5) in a 3' direction, each successive gene displays a more anterior boundary of expression (summarised in Figure 4.7). The cut-off for Hox 2.5 in the spinal cord maps opposite the third cervical prevertebra, and the other six genes have anterior boundaries in the myelencephalon. These findings clearly show that the position of a gene within the Hox 2 cluster is correlated with its relative domain of expression along the anteroposterior axis of the embryo, suggesting that these genes may be playing a role in establishing positional information. This is analogous to the pattern observed for the *Drosophila* BX-C and ANT-C clusters. The physical order of the genes in these clusters reflects the anteroposterior order of the segments each gene affects and their relative domains of expression.

The domains of expression of Hox 2 genes in other ectodermal derivatives mirror their domains in the CNS. In the peripheral nervous system posterior regions express all Hox 2 genes while anterior components only express more 3' members. Thus the dorsal root ganglia, which are derived from trunk neural crest and lie on either side of the spinal cord, express all members of Hox 2 (LeDouarin, 1982). Yet in more rostral regions differential expression of these genes is observed. Hox 2.5 is not expressed in the Xth or IXth cranial nerves, while Hox 2.2 and 2.1 are expressed in the inferior ganglia of the Xth, the nodose ganglion, but not in the inferior ganglion of the IXth nerve (Figure 4.8). Another more 3' gene, Hox 2.7, may be expressed in the inferior ganglia of both the Xth and IXth nerves (Figure 4.8). The inferior ganglia of both the Xth and IXth cranial nerves have been described as originating at a distance laterally from the rhombencephalon and as such more rostral than the sites of Hox 2.5 expression in the central nervous system (Altman and Bayer,
The inferior ganglia of the Xth and IXth nerves have been shown, in the avian and possibly also in the mammalian embryo, to have contributions from both neural crest and ectodermal placodes (Verwoerd and van Oostrom, 1979, D'Amico-Martel and Noden, 1983). The neural cells are derived from the ectodermal placodes while the Schwann cells and satellite cells are formed from neural crest. It is impossible to say from these in situ hybridisation studies whether the expressing cells in the nodose ganglion are of neural crest or placodal origin. None of these genes have been detected in the more rostral superior ganglia of the Xth and IXth nerves which are exclusively derived from neural crest in the avian embryo (Altman and Bayer, 1982, D'Amico-Martel and Noden, 1983). Nor has expression been detected for any of these genes in the more rostral trigeminal (Vth) nerve (Altman and Bayer, 1982, Rugh, 1968, Theiler, 1972).

One can also observe differential expression of Hox 2 genes in the mesenchyme of the thyroid which has a neural crest origin (Le Douarin, 1982). This relatively anterior group of cells do not express the most 5' member of Hox 2, Hox 2.5, but do show expression of the Hox 2.7 which lies at the other end of the cluster (Figure 4.9A). These mesenchymal cells are thought to arise from the hindbrain region, in both avian and mammalian embryos (LeDouarin, 1982). Since Hox 2.5 expression does not extend into the hindbrain while Hox 2.7 does, it is not surprising that these thyroid cells show differential expression of these two genes.

It is suggested that the patterns of Hox 2 expression that are observed in the neural crest and ectodermal placode derivatives correspond with the patterns of expression that have been described in the central nervous system. As such, the described patterns of expression of these genes are consistent with suggested neural crest and ectodermal placode fate maps.

4.7.2 Hox 2 gene expression in the hindbrain and rhombomeres

An important question from our results is why many of the Hox 2 genes have anterior boundaries of expression in the myelencephalon. In early stages of neural development, periodic swellings along the axis of the neural epithelium have been observed and are termed rhombomeres, in the hindbrain. Previously it had not been clear whether segmentation of the neural tube, as defined by rhombomeres, plays any role in establishing the underlying pattern of development in the hindbrain. Recently, Lumsden and Keynes (1989) have presented data that rhombomeres do play a significant role in the segmental patterns of neuronal development in the chick hindbrain. Consistent with these results, Wilkinson et al. (1989a) have shown that rhombomeres represent domains of expression for the zinc finger gene Krox-20, providing molecular...
support for segmentation of the CNS.

It has recently been shown that the anterior boundaries of the more 3' members of Hox 2, Hox 2.6, 2.7, and 2.8 do abut rhombomere boundaries and show a two segment periodicity (Wilkinson et al., 1989b). More curiously the Hox 2.9 gene, which is currently the most 3' member of the Hox 2 cluster, is expressed in the hindbrain wholly within one rhombomere, rhombomere 4 (Murphy et al., 1989, Wilkinson et al., 1989b). Thus it has been suggested that in the hindbrain region homeobox gene do respect segmental boundaries and are acting to confer regional identity on these segments.

One problem with this model is that only Hox 2.6, 2.7, 2.8 and 2.9 respect obvious rhombomeric boundaries while Hox 2.5 has an anterior boundary in the spinal cord and Hox 2.4, 2.3, 2.2 and 2.1 have anterior limits of expression in the unsegmented region of the hindbrain, termed rhombomere 8, r8. While chick only has seven clear rhombomeres lower vertebrates may have more. A study of the zebrafish, Brachidanyo rerio, has suggested that besides the seven that are obvious in higher vertebrates there may also be another three more caudal segments (Hanneman et al., 1988). It is possible that the zebrafish Hox 2 genes (Njolstad et al., 1989) may also respect rhombomeric boundaries and that the more 5' members of the cluster may respect the more caudal segments that are apparent in this organism. While these segments would have been obscured in subsequent vertebrate evolution it is possible that the Hox 2 genes in higher vertebrates are respecting the boundaries that are still morphologically evident in fish.

It is also interesting that Hox 2.5 expresses from most posterior regions to a level in the spinal cord at the third or fourth cervical prevertebrae. This boundary is the level at which the first permanent dorsal root ganglia forms and such the boundary between trunk and occipital regions. Therefore it is possible that Hox 2.5 expression marks the difference between the trunk and the occipital portions of the nerve cord or neural crest. There have been reports of differences between neural crest from these regions such as the inability of occipital neural crest to form sympathetic neurons (Newgreen, 1979, Newgreen et al., 1980). Thus it is possible that the differential expression of Hox 2 genes acts to specify regional diversity in the neural tube. In this model Hox 2.5 would determine trunk and more anterior regions of the neural tube would be sectored through expression of the other genes.

It is important to note that rhombomeres are transient structures and that the anterior limits of the Hox 2 genes remain after the rhombomeric boundaries have disappeared. Rhombomeres are most evident during the period of
motor neuron differentiation and it seems plausible that the Hox genes are acting to confer identity on the segments during motor neuron maturation. But at 12.5 days of gestation after the rhombomeres have disappeared there is still a lot of neurogenesis and the anterior boundaries of the Hox 2 genes are still in the same relative positions. It would still seem likely that at this stage and at later stages the Hox 2 genes are conferring regional cues. Indeed the more 3' genes may act as a molecular remnant of the previous rhombomeric organisation.

This situation is clearly analogous to that of the *Drosophila* homeotic genes most of which do respect segmental boundaries but some of which do not. An example of a gene which does not respect segmental boundaries is that of *lab* which has been described as being expressed in a region specific manner independent of segmentation (Hoey et al., 1986). The unifying theme between *Drosophila* genes of the ANT-C and BX-C and the murine Hox genes is that they would appear to be involved in the specification of regional identity along the rostrocaudal axis and that these gene respect segment boundaries when and where they encounter them.

The Hox 2.9 expression pattern reported by Murphy et al., (1989) and by Wilkinson et al., (1989b) is curious not only because it is restricted to a single rhombomere but also because while it is currently the most 3' member of Hox 2 it is not the most rostrally expressed gene. This gene is expressed caudally of its adjacent 5' gene, Hox 2.8 (Wilkinson et al., 1989b). Therefore it would appear that the absolute correlation between the physical order of the genes and their expression along the rostrocaudal axis does not hold. It will be important to determine if Hox 2.9 is the most 3' member of Hox 2 or not and if any more 3' gene does exist if this gene extends anteriorly of the Hox 2.8 gene.

**4.7.3 Hox 2 gene expression in the mesoderm**

The expression of all genes in posterior regions and only of the more 3' genes in anterior regions is again illustrated in mesoderm. The somitic mesoderm exhibits expression of Hox 2 genes from most posterior prevertebrae to different anterior prevertebrae. Hox 2.2 is expressed up to the eighth prevertebra, Hox 2.1 to the seventh prevertebra while Hox 2.7, the most 3' gene studied, is the most anteriorly expressed extending to the first prevertebra (Figure 4.10). As for the viscera, the metanephric kidney, which originates relatively posterior at the end of the nephrogenic cord adjacent to the cloaca, can be shown to transcribe all Hox 2 members (Saxen et al., 1986) (Figure 4.14). Expression of Hox 2.4, 2.2 and 2.6 has been shown in the mesonephric kidney suggesting that 5' genes in Hox 2 are expressed in the
mesonephric kidney (Figure 4.13). Yet as the mesonephric tubules run for quite a length along the rostrocaudal axis, it would be unfair to say that all Hox 2 genes are expressed in all portions of the mesonephric tubules (Torrey, 1965). Hox 2.5, for example, may not be expressed in more anterior portions of the mesonephric kidney, although this point has not been properly studied.

Differential expression of Hox 2 genes can be seen in the more anteriorly derived developing stomach, lung, gut and gonad (DeHaan and Ursprung, 1965, Balinsky, 1981) (Figures 4.11 & 4.12). One cannot detect Hox 2.5, or 2.4 in the stomach yet transcripts from Hox 2.2, 2.1, 2.6 and 2.7 can be found. This same situation is also true of the lung. Interestingly none of the Hox 2 genes are expressed in the heart which develops in a very anterior region (DeHaan, 1965). Yet in the gonad Hox 2.2 cannot be shown to be expressed along with 2.1, 2.6 and 2.7. While in the gut Hox 2.1 is expressed in the enteric ganglia while 2.6 and 2.7 are expressed in the musculature.

Considering the sites of origin of the mesodermal organs it would seem that the Hox 2 genes are again responding to rostrocaudal position and generally confirm the presumptive sites of origin of the mesodermal tissue. Yet the fact that Hox 2.2 is expressed in the lung but not in the gonad is confusing. It has been suggested that the expression of fewer and more 3' genes of Hox 2 reflects a more anterior position. Therefore from the above results one would suggest that the gonad is derived more anteriorly than the lung mesenchyme. This is contrary to available evidence which locates the developing somatic portion of the gonad at the top of the mesonephros i.e. caudal of the developing lung (Paranko, 1987, Zamboni and Upadhayay, 1982, Yoshinga et al., 1988). This sort of situation is also true of the developing gut which does not express Hox 2.5 but does express Hox 2.6 and 2.7 even though it lies in a caudal region. While the expression patterns are puzzling it is important to note that if a given gene is expressed then the genes 3' of it will also be expressed while those that lie 5' will not. This sort of expression for members of Hox 2 is observed in all other instances, where there is a clearer relationship between the sites of expression and position along the rostrocaudal axis. One does not observe sporadic expression of members of Hox 2.

One problem with this discussion is that it is assumed that the site where the organ is first evident is also the site of origin of its mesodermal cells. This has not been shown to be true for these organs in the mouse. Another factor that one must take into account is time. These results represent the situation at 12.5 and 14.5 days of gestation but at earlier times the Hox 2 genes may display different patterns of expression. It is also
possible that Hox 2 gene expression in the gonad, the lung and the gut are responding to different cues in separate embryonic fields. Thus although it seems that mesodermal structures generally express Hox 2 genes in relation to their position the situation that is observed is not as clear cut as that in ectodermal derivatives. While anteroposterior position is certainly a very important factor in determining Hox 2 gene expression in mesoderm it must be modified by other factors to produce the observed patterns.

4.7.4 The anterior boundaries in mesoderm and in ectoderm are not coincident

In terms of actual position along the anteroposterior axis the anterior limit of expression in mesodermal structures, as compared to ectodermal structures, is more caudal. This is best illustrated by comparing the anterior boundaries of the same gene in the CNS and in the prevertebrae (Figure 4.1). Thus the anterior boundary of expression of a given gene in ectoderm is more rostral than the anterior boundary for that gene in the mesoderm. This may indicate that these genes are under different controls in the different germ layers. This is an attractive idea given the complications that are observed in the patterns of expression of these genes in the visceral mesoderm. In fact this situation has been suggested for the *Drosophila* homeotic genes (Martinez-Arias et al., 1987). Yet as with expression in ectodermal derivatives one finds that all members of Hox 2 are expressed in mesodermal structures that are derived from posterior locations, such as the kidney and posterior prevertebra, and that only the more 3′ members of the locus are expressed in those structures that are more anteriorly derived, such as the lung and the anterior prevertebra.

It is clear that Hox 2 genes are differentially expressed along the rostrocaudal axis of the mouse embryo and that the position of a gene within the cluster is related to its extent of expression along the rostrocaudal axis. While this is most apparent in the central nervous system it is also evident in other ectodermal and mesodermal derivatives.

4.7.5 Correlation between gene order and expression is a general aspect of homeobox clusters

The results described above are in agreement with those of Gaunt et al. (1988) and Duboule and Dolle (1989) who have also shown a similar correlation between the position of some genes within other murine clusters and their relative extent of expression along the rostrocaudal axis. In all four of the murine homeobox clusters, the genes located at the 5′ end are expressed in more posterior domains than those located in the 3′ part of the same cluster. It will be important to determine to what extent murine members of the same
subfamily share similar anteroposterior domains of expression. The Dfd subgroup (Hox 2.6, 1.4, 5.1) appear to have very similar patterns of expression (Gaunt et al. 1989). However in the Abd-B subfamily the Hox 2.5 member is expressed in more anterior regions of the CNS than the Hox 5.2 member (Duboule and Dolle, 1989) and both Hox 2.5 and 2.4 are expressed more anteriorly than Hox 3.1, a member of the Hox 2.4 subfamily (Brier et al., 1988). This suggests that genes in different clusters that are members of the same subfamily do not necessarily have identical domains of expression.

In both mouse and Drosophila, anteroposterior expression of homeobox genes correlates with the organisation of the complexes. But what is even more striking is that the homeobox sequence of the most posteriorly expressed member of Hox 2 (Hox 2.5) is most closely related to the most posteriorly expressed member of BX-C (Abd-B). This correlation holds for the other genes in both the mouse and Drosophila clusters. For example Scr, the Hox 2.1 related gene, is expressed posteriorly of Dfd, the Hox 2.6 related gene and from above it is evident that Hox 2.1 is also expressed posteriorly of Hox 2.6. From this it is concluded that there is a clear relationship between the relative position, sequence identity and domains of expression along the anteroposterior axis for both the mouse and Drosophila homeobox gene complexes. By inference this strongly argues for the existence of an ancestral cluster, prior to the divergence of the protostomes and deuto-stomes, that also exhibited this correlation between the order of the genes and their expression along the rostrocaudal axis. From sequence comparisons it is suggested that this ancient cluster would have had at least six homeobox containing genes and that these genes were also involved in the interpretation of rostrocaudal position. Thus homeobox gene clusters would represent a very ancient and ubiquitous system for the interpretation of rostrocaudal positional information.
5.1 Introduction

Many of the *Drosophila* segmentation and homeotic genes are expressed in the embryonic nervous system. The segmentation genes *eve*, *en*, and *ftz* are expressed in a specific subset of neurons in every segment of the developing central nervous system and seem to be involved in neuronal determination (Doe et al., 1988b, 1988c). The homeotic genes are expressed at their highest levels in the embryonic nervous system (Doe and Scott, 1988a). Unlike the segmentation genes, the homeotic genes are not expressed in every segment of the CNS and as such are not thought to be involved in the specification of segmentally reiterated features of the CNS. Instead, these genes tend to show peak expression in given parasegments with lower levels of expression in more posterior regions and are thought to regulate correct anteroposterior differentiation in the CNS (Doe and Scott, 1988a). Mutations in these genes affect those regions that exhibit the highest levels of expression of the genes.

The expression patterns of the murine homeobox genes along the length of the CNS are remarkably similar to the situation in *Drosophila*, with strong anterior boundaries and expression trailing back through more posterior regions. Unlike *Drosophila*, the role of the murine homeobox genes in the ontogeny of the central nervous system is not clear.

The mammalian CNS lies dorsally is considerably more complex than the ventral *Drosophila* CNS. The mammalian CNS develops from a pseudostratified neuroepithelium and over a period generates the diverse range of cell types. The transverse organisation of the mammalian spinal cord and hindbrain is relatively similar throughout its length rostral to caudal, except for differences in timing of maturation of given classes of neurons. The different neuronal cell types are organised and show spatial and temporal development in the dorsoventral plane (Altman and Bayer 1984). The large motor neurons of the ventral horn are the first major group of neurons that are born and among the earliest maturing cells of the spinal cord. This is followed by the relay neurons which are located laterally and lastly by the interneurons of the dorsal horn which are among the latest born. While the motor neurons are the first major class of neurons to be born, their development is earlier and more advanced in rostral regions as compared to caudal regions. This also tends to be true of the other neurons. Thus the development of the CNS occurs with rough ventrodorsal and rostrocaudal temporal gradients, such that the first
major class of neurons are born ventrally and maturation is more advanced in rostral regions as compared to caudal regions.

If one is to approach the role of murine homeobox genes in the ontogeny of the CNS it is important to analyse their expression patterns, temporally and spatially, in the transverse plane. A number of groups have analysed the expression patterns of these genes at stages in the development of the CNS and uniform, dorsal and ventral patterns of expression have been described (Brier et al., 1988, Dony and Gruss, 1987, Toth et al., 1987, Graham et al., 1988, Duboule and Dolle, 1989, Krumlauf et al., 1987). There has also been one study of the temporal patterns of expression of a single gene, Hox 2.5, which was shown to display different patterns at different time points (Bogarad et al., 1989). The fact that one observes changing patterns in the expression of one gene underlines the importance of analysing expression patterns at several time points.

In this study the expression patterns of seven members of the Hox 2 cluster have been analysed during development of the CNS. Expression of all of these genes have been analysed on adjacent sections at multiple positions along the anteroposterior axis at several time points. These genes all show the same dynamic patterns of expression and only show differences along the anteroposterior axis, as has been illustrated previously (Graham et al., 1989). These rostrocaudal differences were used to suggest that these genes are involved in patterning along this axis. Yet the fact that there are no dorsoventral differences between any of these genes would suggest that they are responding to events that are occurring dorsoventral in the nervous system and are not actually involved in dorsoventral patterning. It is suggested that these genes are expressed transiently to confer anteroposterior position on newly born neuronal cells. As such one would expect these genes to show patterns of expression that mirror cell maturation in the dorsoventral axis.

5.2 Hox 2 genes are dorsally restricted at 12.5 days

The isolation, and use of, gene specific probes for 7 members of Hox 2, Hox 2.1 to 2.7, has been previously described (See chapter 4). In this study these probes were used in in situ hybridisation experiments to compare the patterns of expression of these genes in serial transverse sections of the nervous system. Embryos from different stages were analysed at several points along the anteroposterior axis.

Since 12.5 days p.c. is the stage at which these genes show their highest levels of expression this time point has often been used as a convenient stage for the analysis of the patterns of expression of Hox 2 genes, the
Figure 5.1 - Expression patterns of four members of Hox 2 at five different points along the rostrocaudal axis of a 12.5 day mouse embryo. Sections were taken at five different points labelled A-E. The rostral to caudal order of the sections is indicated at the top of the figure. Representative bright field photographs are shown for each plane of section. The dark ground photographs display the patterns of expression of each gene at each of the five different points. The expression pattern changes from an "M" shape caudally to a sharp dorsal domain of expression in rostral regions. The probe used for each row of sections is marked on the left, the position of the section is marked at the top of each column. B - A bright field photograph of a 12.5 day mouse embryo is used to illustrate the plane of each set of sections, A-E. The organisation of the Hox 2 cluster is shown at the bottom. The number in parenthesis indicates the chromosomal location of this cluster. The genes are expressed in order from posterior to anterior as is indicated above the cluster.
organisation of which is again illustrated on figure 5.1 B. Figure 5.1 A shows the patterns of expression of four members of Hox 2 on adjacent transverse sections at five different points along the rostrocaudal axis, labelled A-E on figure 5.1 B. One can easily see that Hox 2.5, 2.4 2.2 and 2.7 all display similar patterns of expression at this stage. In the sections from position A, figure 5.1 A a,f,k,p, are caudal and cut through the metanephric kidney. One can note that all of the genes exhibit an "M" shaped pattern with strong lateral/dorsal labelling. As one moves rostrally along the axis of the animal the "M" shape is still evident up to the level of the lungs, at B, sections b,g,l,q, in figure 5.1 A. More rostrally still the pattern changes such that all genes exhibit a strong dorsal domain of expression with a sharp restriction. This boundary in the expression pattern is particularly interesting as it does not seem to correlate with any obvious morphological boundary. This is evident in sections from positions C and D of figure 5.1. In all cases there is generally lower or absent expression in the ventricular layer, that layer which lies immediately adjacent to the lumen.

In the hindbrain region, E in figures 5.1 A and B, there is no expression of Hox 2.5 but the other three genes, 2.2, 2.2, and 2.7 are still expressed. The observed difference between the expression of Hox 2.5 and that of the other three genes in the hindbrain results from the fact that there is a correlation between the physical order of genes in Hox 2 and their expression along the rostrocaudal axis (see chapter 4, Graham et al., 1989). This has also been shown to be true of the other murine Hox clusters (Gaunt et al., 1988, Duboule and Dolle, 1989). The Hox 2 cluster is illustrated in figure 5.1 B and it is clear that as one moves along the cluster in a 5' to 3' direction each successive gene has a more anterior boundary of expression. Thus Hox 2.5 is not expressed in the hindbrain, figure 5.1 A e, since its rostral extent of expression maps in the spinal cord. It is important to note that the only differences that one observes between the genes in a transverse analysis of expression in the central nervous system is that imposed by the rostrocaudal axis. Otherwise the genes show the same patterns of expression at the same positions. It is interesting that the pattern that one observes changes along the rostro-caudal axis from an "M" shape to a sharp dorsal restriction.

To address how these patterns were generated the expression of these seven Hox 2 genes were analysed at both earlier and later stages during the development of central nervous system. Again all of the seven Hox 2 genes, that were analysed, show identical patterns of expression.
Figure 5.2 - The patterns of expression for seven members of the Hox 2 cluster in the 10.5 day nerve cord. Transverse sections have been taken at various points along the rostrocaudal axis and have been probed with each of the seven probes and the results are shown in dark ground. Each gene is expressed across the nerve cord. The probes were as follows: a- Hox 2.5, b- Hox 2.3, c- Hox 2.4, d- Hox 2.2, e- Hox 2.1, f & h- Hox 2.6, g & i- Hox 2.7.
Figure 5.3 - Change in expression pattern of Hox 2.2 in the spinal cord between 10.5 and 11.5 days of gestation. Transverse sections have been taken from the spinal cord of a 10.5 and 11.5 day mouse embryo, shown in bright field on the left, and probed with the Hox 2.2 gene. The *in situ* results are displayed in dark ground on the right. DRG- Dorsal Root Ganglia, M- Motor Horn
5.3 Hox genes are expressed uniformly across the nerve cord at 10.5 days

There is clearly a temporal aspect to the patterns of expression that one finds for the Hox 2 genes in the developing central nervous system since a dorsal restriction in the expression of these genes at earlier stages is not evident. Expression of all of the genes at 10.5 days is shown in figure 5.2 and one can observe that all genes are uniformly expressed across the neural tube at this stage. This uniform pattern of expression is observed at all points along the axis that are within the rostrocaudal domain of expression of each gene. The sections in figure 5.2 are not all adjacent but are from three different positions along the rostrocaudal axis. Sections a and b are from the hind limb level and have been probed with Hox 2.5 and 2.3 respectively while sections c and d are from the prospective lung region and have been hybridised with Hox 2.4 and 2.2 probes. Sections e, f and g are adjacent hindbrain sections that illustrate the expression of Hox 2.1, 2.6 and 2.7, respectively. The strong expression of Hox 2.6 and 2.7 across the hindbrain is also displayed in h and i. From these results it is clear that all of the Hox 2 genes analysed are expressed uniformly across the neural tube at 10.5 days of gestation.

10.5 days of embryogenesis is the period during which the dorsal root ganglia are formed (Rugh, 1968, Wentworth, 1984b). These sensory ganglia are derived from neural crest and are metamerically organised on either side of the spinal cord. In figure 2 one can observe the expression of Hox 2.5, 2.4, 2.3 and 2.2 in the forming dorsal root ganglia. Hox 2.1, 2.6 and 2.7 are also expressed in these structures although this is not shown in figure 5.2 since the sections that have been probed for these genes lies more rostral than the extent of these ganglia.

5.4 Strong lateral expression of Hox 2 genes at 11.5 days

This pattern changes at 11.5 days. This is shown for the Hox 2.2 gene in figure 5.3. While the expression at 10.5 days of gestation is clearly all across the nerve cord a day later expression is no longer spread evenly dorsal to ventral but is now abundant in lateral regions, down graded in motor horns and still generally expressed in other regions. One can also note the expression in the forming dorsal root ganglia (Figure 5.3). Figure 5.4 shows expression of six members of Hox 2 at various points along the anteroposterior axis of the central nervous system. Again the patterns displayed by the genes are identical at other points along this axis (data not shown). Sections a and d are from relatively caudal regions while b, c, e, and f are from the hindbrain region. It is clear from figure 5.4 that for all of the genes there is stronger expression, although not exclusive, in the lateral
Figure 5.4 - Transverse sections of the 11.5 day mouse nerve cord probed with Hox 2 genes. The results showing strong lateral expression are displayed in dark ground. The probes used were as follows: a- Hox 2.5, b- Hox 2.4, c- Hox 2.2, d- Hox 2.1, e- Hox 2.6, f- Hox 2.7.
Figure 5.5 - Expression of three Hox 2 genes at three different points along the rostrocaudal axis of the 14.5 day mouse embryo. Each column of sections have been probed with the same gene as is indicated at the base. Each row of sections is from the same position along the rostrocaudal axis. The results reveal a reappearance of ventral expression in more rostral regions and are shown in dark ground.
regions. At this stage the motor horns are visible and do not show as high levels of Hox 2 transcripts. This is most obvious for Hox 2.2 in figure 5.4d.

5.5 Expression of Hox 2 genes reappear ventrally at 14.5 days

Analysis of later stage of development reveals that the pattern described for the rostral 12.5 day cord is also observed caudally at 14.5 days. This is illustrated very clearly for Hox 2.5, 2.4 and 2.2 in figure 5.5 particularly in sections c,f, and i which cut through the metanephric kidney and the more rostral sections b,e and h which are at the lung level. Again all of the Hox 2 genes are dorsally restricted and a sharp demarcation, that does not correlate with any morphological boundary, is revealed. This pattern changes as one move more anterior with expression turning on again in the ventral half of the neural tube. This is shown for Hox 2.4 and 2.2 in figure 5.5 on sections d and g. Hox 2.5 does not exhibit this pattern since this pattern is only evident in regions which are more anterior than the boundary of Hox 2.5 expression. Thus again the only difference that one observes between these genes relates to their differential expression along the anteroposterior axis. This difference in expression is not restricted to ectodermal tissues but is also found in mesodermal derivatives. One can note the expression of Hox 2.2 in the mesenchyme of the lung in figure 5.5 section h but not the expression of Hox 2.5 or 2.4 in the adjacent sections, b and e.

5.6 Hox 3.1 is expressed in stripes

The sharp dorsal restriction in expression is particularly interesting. This pattern was compared with that of a homeobox gene from another cluster. The Hox 3.1 gene was chosen since this gene had previously been described as exhibiting a ventral domain of expression in the 12.5 day nerve cord although without a sharp restriction (Breier et al., 1988). Adjacent sections were probed with Hox 2.5 and Hox 3.1 and the expression in the neural tube analysed. In rostral regions Hox 2.5 displayed a sharp dorsal domain of expression while Hox 3.1, as previously reported, was most abundant ventrally and did not display a sharp demarcation in expression. Yet more interestingly in caudal regions while Hox 2.5 displayed the typical "M" shape pattern Hox 3.1 exhibited a strikingly different pattern. This pattern is shown in figure 5.6. Hox 3.1 is expressed in a stripe across each half of the neural tube projecting laterally from the edge of the ventricular zone. There is also expression of this gene along the sides of the ventricular layer. Expression is absent from other regions. Figure 5.6 shows the expression of Hox 2.5 on an adjacent section displaying the previously described "M" shaped pattern. Thus the expression pattern of at least two genes from different clusters are radically different at the same stage of development.
Figure 5.6 - A comparison between the expression patterns of the Hox 2.5 and the Hox 3.1 genes in the caudal 12.5 day mouse spinal cord. The bright field photographs show the regions that were probed and the very different hybridisation patterns that were revealed are shown in dark ground in the lower two rows in lower and higher power. While Hox 2.5 is expressed dorsally Hox 3.1 is expressed in stripes. The probe used is indicated at the base of each column.
5.7 Discussion

The expression patterns, in transverse sections of the nerve cord, for seven members of Hox 2, Hox 2.1 to 2.7, have been compared on serial sections. This comparison has used embryos at different stages and sections have been taken from numerous points along the rostrocaudal axis. The results show dynamic patterns of expression for these genes during the development of the central nervous system. It is important to note that all seven genes display essentially identical patterns of expression. The only differences that one can observe between any of these genes at any of the stages analysed, from 10.5 - 14.5 days of gestation, is in their expression along the rostrocaudal axis. As has been previously reported the genes of the Hox 2 cluster show differential expression along the rostrocaudal axis such that there is a correlation between the physical order of the genes and their expression along this axis. As one moves from Hox 2.5 in a 3' direction each successive gene shows a progressively more anterior limit of expression (Graham et al., 1989).

From the earliest times analysed (9 days of gestation) these genes show expression across transverse sections of the neural tube and show expression in the closing neural folds (data not shown). At 10.5 days of gestation the expression of these genes can still be seen to be all across transverse sections of the neural tube. A day later members of this cluster show strong lateral expression, nil or very low levels in the motor horns and general expression in other regions of the nerve cord. By 12.5 days the caudal regions of the spinal cord exhibit a more dorsalised pattern with expression being found in lateral dorsal regions but not in the ventricular zone or in ventral areas, giving an "M" shape. Yet in rostral portions of the 12.5 day CNS the transcripts of these genes show dorsal restriction with a sharp demarcation, that does not correspond to any obvious morphological boundary. Dorsal restriction has also been described for the Hox 1.4 gene and again at 12.5 days this gene was found not to be expressed in the ventricular zone (Toth et al. 1987). The rostral 12.5 day pattern can still be observed in caudal portions of the 14.5 day spinal cord but not in more rostral regions. In rostral regions of the 14.5 day CNS do not display this sharp dorsal restriction but expression can be found across the neural tube in both ventral and dorsal regions.

These results are consistent with those of earlier studies on the expression patterns of members of Hox 2 and in particular confirm the results of Bogarad et al., (1989) with the Hox 2.5 gene who describe identical spatial and temporal patterns of expression. The pattern that was previously described
for Hox 2.6 with expression across the cord is thought to represent an intermediate pattern between that observed at 11.5 days and that at 12.5 days (Graham et al., 1988). One problem in using mice of the CBA strain is the lack of synchrony in any given litter, thus it is difficult to give accurate estimations based on timings of the stage of a given embryo. The expression of Ghox 2.1 in the CNS, the chicken homologue of Hox 2.1, has been shown to be predominantly dorsally located at stage 25, a stage that is approximately equivalent to a 12 to 13 day mouse embryo (Hamburger and Hamilton, 1951, Wedden et al., 1989). Thus in two vertebrate species two homologous genes show homologous patterns of expression in the CNS. This may suggest that these genes are performing identical functions in both species.

The Hox 3.1 pattern of expression is obviously very different from that of members of Hox 2. Not only is it ventrally abundant in the rostral portion of the nerve cord at 12.5 days of gestation, while the Hox 2 genes display a sharp dorsal domain of expression, but in caudal regions Hox 3.1 is expressed in a stripe across each side of the spinal cord extending laterally from the edge of the ventricular zone. At this level it is also expressed along the side of the ventricular zone. A study of earlier patterns of expression of Hox 3.1 has not been conducted and it will be important to analyse these earlier stages if we are to understand the observed 12.5 day patterns.

It is clear that the Hox 2 genes show differential rostrocaudal expression and this fact has been used to suggest that they may play a role in rostrocaudal patterning, yet these same genes show no difference in the transverse plane of the spinal cord. Since none of these genes show any differences in their transverse patterns of expression in the developing spinal cord it is suggested that the dynamic patterns that are observed reflect events in the spinal cord. If these genes were involved in dorsoventral patterning of the spinal cord one would expect to observe difference between the genes.

5.7.1 Hox 2 gene expression correlates with the time and place of the birth of the major classes of spinal neurons

An important question now emerges as to what developmental events these genes are following? It would appear not to be something as simple as proliferation of cells since short term thymidine labelling, that would mark cells in S-phase, of mouse neural tubes at various stages does not produce the same patterns, at comparable stages, as are observed with the Hox 2 probes (Nornes and Carry 1978). Nor does it seem likely that these genes are involved in the basic neuronal differentiation since neurons of the same class located at different points along the rostrocaudal axis do not express the same genes.
The dynamic patterns of Hox 2 expression are more similar, in both temporal and spatial aspects, to that observed for the birth of the major classes of spinal neurons (Sims and Vaughn, 1979, Altman and Bayer, 1984, Wentworth, 1984a, 1984b).

The first major class of neurons to differentiate are the motor neurons, which arise ventrally at about day 9 to 11. The next major class are the commissural relay neurons whose cell bodies have been described as lying in lateral regions from ventral to dorsal. The early differentiating commissural neurons originate between 10.5 and 12.5 days p.c. Association or funicular relay neurons are born just behind the commissural neurons and have cell bodies located dorsal of the oval bundle of His and as far ventral as the lateral funiculus (Wentworth, 1984b). Lastly the inter neurons of the dorsal horn are born between 12 and 14 days p.c.

It is suggested that the Hox 2 genes are expressed across the early neural tube and as neurons are born they transiently express Hox 2 genes at high levels. The transient expression of these genes is to allow the differentiating neuron to understand its position along the rostrocaudal axis, through which Hox 2 genes it expresses. Once the cells have begun to be fully differentiated they already know their position and turn off expression of these genes.

This scenario would suggest that the early pattern of expression across transverse sections of the neural tube is temporally and spatially modified as each of the major classes of neuron are born. Firstly one would expect to see expression in the ventrally lying motor neurons and then these ventral cells would turn off and expression would be high in the forming commissural neurons whose cells bodies should lie lateral at about 11.5 days. Expression would then recede dorsally, but still be lateral, as the next major class, the association neurons, are born. Finally one would expect to see general dorsal expression as the inter neurons of the dorsal horn are being born between 12 and 14 days. Since the neural tube matures in a rostral to caudal direction we would suggest that the differences between caudal and rostral portions at 12.5 days reflects difference in development along this axis. This suggestion is also supported by the fact that the rostral 12.5 day pattern is still observed in caudal portions of older embryos. It would moreover explain the differences observed at 14.5 days of gestation along the rostrocaudal axis.

5.7.2 The Hox 2 genes and the Hox 3.1 gene must be responding to different signals

Thus the role of the Hox 2 genes would be to allow cells to understand their rostrocaudal position and the dynamic patterns that are observed in the
neural tube result from the birth of the different classes of neuron. This would seem to be the most plausible association between the observed patterns and events during ontogeny of the central nervous system. Although from the Hox 3.1 pattern that has been described it would seem that this is not the case, for this gene at least. The expression pattern displayed by this gene does not fit with the above interpretation of Hox 2 expression pattern and it seems likely that this gene must be responding to different cues than those that organise the Hox 2 expression.

It has been demonstrated that genes that share similar positions within different Hox clusters exhibit high degrees of sequence identity and it has been suggested that these gene clusters all arose from a common ancestral cluster (see chapters 3 and 4, Duboule and Dolle, 1989, Graham et al., 1989). Related genes within different clusters have been shown to display identical rostral limits of expression in the CNS (Gaunt et al., 1989). If, as has been proposed, these genes are involved in the establishment of rostrocaudal positional information it would seem fruitless to utilise many Hox genes to specify the same position. Yet the difference between the expression in the CNS of Hox 3.1 and Hox 2.4, its closest Hox 2 gene, may point out that while genes have the same anterior boundary in expression they are displaying very different patterns of expression in the dorsoventral plane. Thus it is possible that during deutrostome evolution as the central nervous system became more complex the Hox gene clusters were duplicated to increase the potential to specify different positions. It will obviously be important to determine the expression patterns of the genes from the other Hox clusters.

5.7.3 Re-expression ventrally and the sharp dorsal demarcation in the expression patterns

There are still some unexplained aspects of the patterns of Hox 2 gene expression. Firstly if the expression of these genes does correlate with aspects of neurogenesis why are these genes re-expressed ventrally in the rostral 14.5 day spinal cord? Possibly the ventral expression reflects aspects of gliogenesis, although this is uncertain. Another major point is the sharp line that demarks the dorsal expression of these genes at 12.5 and 14.5 days of gestation.

As has been said this line does not correspond with any obvious morphological marker that is evident at this stage yet the transient nature of our line, which has been revealed by at least seven different homeobox genes would suggest that it is of significance. One morphological marker that would lie in a similar position and divide alar and basal plates is the sulcus limitans that has been described in rat (Altman and Bayer, 1984) and human
(Muller and O’Rahilly, 1988) embryos. Yet this structure has been noted at comparatively earlier stages in rat and human and would seem not to be obvious at 12.5 days in the mouse and later. A similar restriction in the murine spinal cord has also been observed with immunohistochemistry for the cellular retinol binding protein (CRBP) which has revealed a ventral pattern with a sharp line (Maden et al., 1989). This CRBP pattern is established at about 10.5 days of gestation and it is interesting that CRBP respects a line in the neural tube 2 days before the Hox 2 genes. Comparative studies between CRBP and Hox 2 probes reveals that these probes respect the same line (A.Graham, R.Krumlauf and M.Maden unpublished). We have not as yet determined whether the Hox 3.1 line that is apparent in caudal regions of the 12.5 day nerve cord lies along the boundary of CRBP expression. The fact that a transient line that separates alar and basal plates at mid-gestation times is revealed by homeobox gene expression and the presence of retinoid binding protein would suggest that it is of biological/developmental importance. One would predict that other molecules will also exhibit similar types of restriction in the developing nerve cord.

It is intriguing that the line should be revealed by homeobox and retinoid probes since there would appear to be an association between the expression of homeobox genes and retinoic acid such that most members of Hox 2 are activated upon addition of retinoic acid to F9 teratocarcinoma cells (N.Papalopulu pers. comm.). It should also be noted that the line that is evident through CRBP staining appears prior to the Hox 3.1 line which is prior to the sharp dorsal demarcation of Hox 2 genes. This is interesting since in tissue culture the Hox 2 genes respond to the addition of retinoic acid and in the embryonic CNS the retinol/retinoic acid is acting before the homeobox genes. Again it is of interest that the cellular retinoic acid binding protein (CRABP) can be detected at high levels in the forming commissural and association neurons and in neural crest derivatives which also express Hox 2 genes (Holland and Hogan, 1988a, Graham et al., 1988, Maden et al., 1989).

While this study actually focuses on the dynamic spatial and temporal patterns of Hox 2 genes in the transverse plane of the spinal cord it serves to reinforce the importance of their differential expression along the rostrocaudal axis. Hox 2 genes in this system would seem to be acting to confer rostrocaudal position on newly forming neurons, when they need to understand their position. The expression pattern of Hox 3.1 is clearly different and it would seem to be responding to different developmental cues. Yet the roles of murine homeobox genes can be seen to be strikingly similar to
their Drosophila counterparts. In both organisms these genes can be seen to be involved in the specification of anteroposterior position. Whether the murine and Drosophila central nervous systems are actually homologous is in question but it is clear that these genes which are derived from a common ancestor are fulfilling very similar roles i.e. conferring anteroposterior position.
CHAPTER SIX - GENERAL DISCUSSION

6.1 Positions of genes in a cluster reflect domains of expression in the rostrocaudal axis for both mouse and Drosophila

This thesis commenced with a discussion of regional specification (Wolpert, 1971). The work presented in this thesis is consistent with, and strengthens, previous suggestions that murine homeobox genes play a role in the establishment of regional specification (see Holland and Hogan, 1988b). Evidence includes both expression studies and evolutionary considerations.

It has been shown that there is a correlation between the physical order of the members of Hox 2 along the chromosome and their order of expression along the rostrocaudal axis, such that each successive 3' gene is expressed more rostrally (see chapter 4). The differential expression of the Hox 2 genes along the rostrocaudal axis is apparent in both ectodermal and mesodermal derivatives. This ordered pattern of expression would be consistent with these genes having a role in regional specification. This is supported by comparison with the Drosophila homeotic genes of the ANT-C and BX-C. It was for these genes that a correlation between the physical order of the genes and their expression along the anteroposterior axis was first described (Harding et al., 1985, Akam, 1987). As has been discussed in chapter one it was clear that these genes are involved in the establishment of regional specification along the Drosophila rostrocaudal axis (Harding et al., 1985, Akam, 1987). The relationship between the murine and Drosophila clusters extends further than aspects of their expression patterns. Sequence comparisons indicate that the murine Hox clusters and the Drosophila homeobox clusters are both derived from a common ancestral cluster (Chapter 3, Duboule and Dolle, 1989, Graham et al., 1989). The clustering and the ordered expression of the homeobox genes is therefore homologous in both Drosophila and mice and it seems very likely that the murine genes, in common with the Drosophila genes, are acting to confer rostrocaudal regional specification.

Since mice and Drosophila are representatives of the deutostome and protostome lineages, respectively, it is suggested that the ancestral Hox cluster existed prior to the divergence of these two lineages. By inference it is proposed that the correlation between the order of the genes and their expression along the rostrocaudal axis may be a general feature of all higher metazoans. It is also suggested that the Hox clusters are involved in regional specification in all higher metazoans.
6.2 Evolution of homeobox gene clusters and the evolution of the body plan

Sequence comparisons have been used to suggest that the ancestral cluster contained 6 homeobox genes that subsequently expanded during evolution to yield the *Drosophila* and mouse clusters (Figure 3.9, Graham et al., 1989). If these genes are involved in regional specification one would be tempted to link an increase in the number of these genes with an increase in the complexity of the body plan.

The common ancestor to protostomes and deutozystomes is thought to have been acoelornate, bilateral and non-metameric, possibly of a comparable grade of organisation to that of a Plathyhelminthe or a Nemertean (Clark, 1964, Akam, 1989). At this stage in evolution there must already have been molecular representations of back, front and of subdivisions of the middle region of the body. Interestingly during the evolution of the relatively complex organisation of *Drosophila* from that of the putative ancestor it is felt that the cluster of homeobox genes only expanded by two, with the evolution of *abd-A* and *Ubx* (Figures 3.8 and 3.9, Akam, 1989). It is these genes that are responsible for distinguishing between the thoracic and abdominal trunk segments, a distinction which is itself thought to be a relatively recent event in arthropod evolution (Akam, 1989). It would seem that the ends of an animal had been fixed early in evolution and that the duplication of these genes is to cater for internal specialisation. It is also evident from this that one cannot link the evolution of homeotic genes to the more drastic changes in the body plan that must have occurred during the evolution of the diptera (Akam, 1989).

Unlike *Drosophila*, during the evolution of mammals from the common ancestor of the protostomes and deutozystomes the homeobox gene cluster not only expanded but it also duplicated and diverged to give at least four separate but related clusters (Figure 3.8). It would be interesting to understand how morphological changes that occurred during the evolution of the deutozystomes could possibly be related to the increased homeobox gene/cluster diversity. It is likely not to be due to the formation of a coelom as such, since the protostomes also evolved a coelom but did not duplicate the homeobox gene clusters. However, since it is felt that the coelom evolved separately in the protostome and deutozystome lineages the duplication of the Hox cluster(s) could be partially due to both the evolution of, and to the type of coelom that arose in the deutozystome lineage (Clark, 1964). There have been other major changes to the body plan that could be linked to the gene evolution such as the acquisition of a notochord and repeated blocks of muscle along the rostrocaudal axis. Mesodermal segmentation in this lineage could account for
some of the duplication events, as could the subsequent evolution of the vertebral column or the increased complexity of the viscera. The actual order of events in the evolution of the deuterostomes is obscure and even less is known about the temporal sequence of homeobox cluster duplication and diversity, therefore at present it would be very difficult to tie the genetic events with the morphological events.

From figure 3.8 it is clear that not all of the genes or subfamilies are represented in all the murine clusters. Thus although some clusters such as Hox 1 and Hox 2 have expanded from the ancestral cluster, the Hox 3 and Hox 5 clusters appear to be missing some subfamily members that are represented in Hox 1 and Hox 2. The genes that are absent are internal and thought to be derived from Antp and therefore evolutionarily more recent.

Considering these points it is important to note that it has been proposed by Ohno (1970) that during chordate evolution there have thought to have been two whole genome duplications. The first of these is suggested to have occurred between the transition from urochordates to cephalochordates and the second during the evolution of mammals from fish. Since it is suggested that the four murine homeobox clusters arose from a single cluster during deuterostome evolution this has obvious relevance to homeobox gene clusters. It should also be noted that in Brachydanio rerio a Hox 2 and a Hox 3 cluster have been described but not a Hox 5 or a Hox 1 cluster. The Hox 1 and Hox 2 clusters are very close to each other as are the Hox 3 and Hox 5 clusters and therefore a complete genome duplication between fish and mammals could have resulted in the current four homeobox clusters. If this model is true it would also imply that there was a single homeobox cluster in urochordates but due to genome duplication there would be two in cephalochordates. It would be very interesting to extend the studies presented in this thesis to a phylogenetic study of the organisation and the expression pattern of the Hox genes. Hopefully one could begin to understand when the Hox clusters were duplicated and in what sort of animal such an event may have occurred. It may then be possible to relate the evolution of the Hox genes to that of the body plan and to prove or disprove the above proposed theory of genome evolution as it relates to homeobox genes. For example a comparison amongst the various organisms that one finds within the protochordates could be very informative. A comparison between a cephalochordate, such as Amphioxus, which has a full notochord extending from its tail to the tip of the head and has segmental muscle although not a vertebral column, with an ascidian, which does not have segmented muscle, although it has longitudinal muscle fibres, nor does it have as full extended notochord, could
be very informative (Romer, 1976).

6.3 Relationships between the expression patterns of genes from related murine Hox clusters

If these genes are involved in regional specification one would expect that the expansion of these genes/clusters would increase the number of instructions or degrees of freedom available to an embryo. Aspects of this should be revealed by comparing the expression patterns of subfamily members.

An analysis of three murine genes of the same subfamily, Hox 1.4, 2.6 and 5.1, has shown these genes to have the same anterior boundaries in the CNS and in the somitic mesoderm (Gaunt et al., 1989). This study also highlighted important differences between the stage- and tissue-dependent expression of these genes. For example Hox 5.1 is relatively abundant within ectoderm and mesoderm at early stages (8.5 - 9.5 days p.c.), yet was only weakly detected in the mesodermal components of the lung and stomach at 10.5 days and was absent at 12.5 days. In contrast Hox 1.4 and 2.6 transcripts were relatively weak at 8.5 - 9.5 days p.c. but were abundant at 12.5 days. These three genes also show differential expression in retinoic acid induced differentiation of F9 teratocarcinoma cells (N.Papalopulu pers. comm.).

Other subfamilies show even more drastic differences between their members as for example between Hox 2.5 and Hox 5.2. While Hox 2.5 has an anterior boundary at the level of the first permanent dorsal root ganglia, the Hox 5.2 boundary is located very much more posteriorly. Dolle and Duboule (1989) describe very strong expression of Hox 5.2 in the developing forelimb while Hox 2.5 cannot be detected in this structure, at any stage of development. Another example of drastic differences in the expression patterns of subfamily members is that described for Hox 2.4 and Hox 3.1 in the developing nervous system (see chapter 5). While Hox 2.4 exhibits an "M" shaped transverse profile of dorsal expression in the caudal 12.5 day spinal cord the Hox 3.1 gene is expressed in lateral stripes across the neural tube and along the side of the ventricular zone.

The expression patterns of related genes from different subfamilies give no obvious clues as to the possible changes in the body plan that accompanied the duplication and divergence of the Hox clusters. That having been said the differences that are observed in the spinal cord in transverse sections are particularly striking and could well reflect different roles for each cluster in the generation of the complex organisation of the mammalian central nervous system. Thus the evolution of a complex central nervous system could have been associated with the evolution of more Hox gene clusters and with the
The fact that there are similarities in the expression patterns between subfamily members might indicate that these genes share some overlap and redundancy in function, and the differences observed could also usefully increase the variety of instructions available to the embryo. These differences would not necessarily need to be as gross as whole tissue or stage differences in expression. Since these genes are thought to encode transcription factors, it is possible that by varying the levels of these factors one could change the developmental decisions that a cell could make.

6.4 Hox 2 gene expression and the vertebrate head

It is intriguing that the expression of members of Hox 2, or of any other murine Hox genes, in the ectoderm does not extend rostral of the hindbrain. This may suggest that there are more 3’ members of Hox 2, as yet undiscovered, which would be expressed in these more rostral regions. Given that the murine and the Drosophila clusters both evolved from a common ancestral cluster that is thought to have had a lab like gene it is curious that while lab is expressed anteriorly of its adjacent telomeric gene pb, the Hox 2 gene that is most related to lab, Hox 2.9, is expressed caudally of the adjacent 5’ gene Hox 2.8 (Pultz et al., 1988, Diedrich et al., 1989, Wilkinson et al., 1989b). Therefore it would be tempting to suggest that there is at least one other member of Hox 2 that is also related to lab and that this gene maintains the ordered expression. While this may or may not be the case it would still seem unlikely, taking into account the above evolutionary arguments, that there are enough extra members of Hox 2 to provide regional information all the way from the hindbrain to the tip of the forebrain. Contrastingly the Drosophila lab gene is expressed in very anterior regions of the embryo (Hoey et al., 1986, Mlodzik et al., 1988, Diedrich et al., 1989).

The vertebrate head shows numerous differences from the rest of the body particularly in its embryological origin. Unlike the rest of the body, in the head the neural crest functions as mesoderm and forms connective, skeletal and tissue (Le Douarin, 1982). It has been suggested that, in the transition from protochordates to vertebrates and in the switch to an active mode of predation, many features occurring only in vertebrates were concentrated in the head (Gans and Northcutt, 1983). Thus it was necessary to add this "new head" on to the protochordate body and the onus for building much of this structure fell to the neural crest cells (Gans and Northcutt, 1983). The ends of homeobox gene clusters, and therefore the most anteriorly expressing gene, would have been defined prior to this modification. Therefore
the fact that Hox 2 genes do not extend beyond the hindbrain may suggest that these genes still respect the anterior end of the "old body" prior to the addition of the vertebrate head.

To return to phylogenetic approaches it would obviously be of interest to know if a lab-like gene existed in an animal such as Amphioxus, a cephalochordate, which is felt to be close to the sort of animal that existed prior to the vertebrates. If such a gene did exist it would be interesting to determine its extent of expression along the rostrocaudal axis. One would also expect that there would be more recently evolved genes that are used to interpret position in more anterior regions. In light of this it is interesting that sea urchins possess one Engrailed like gene while mice have two and that the murine En-2 gene is expressed in the midbrain (Dolecki and Humphreys, 1988, Joyner and Martin, 1987, Davis et al., 1988). Thus the two engrailed genes were duplicated in the deuterostome lineage, possibly to cope with the vertebrate head.

6.5 Hox genes are involved in the interpretation of regional cues

While this work points towards homeobox genes having a role in the establishment of rostrocaudal regional specification in the mouse this role is probably in the interpretation of information rather than being the primary source of the information. This is clearly homologous to the role played by the Drosophila homeobox gene clusters (see introduction, Akam, 1987, Ingham, 1988). It is also important to note that the homeodomain proteins are nuclear proteins (Harvey et al., 1986, Kessel et al., 1987, Odenwald, et al., 1987). Unlike Drosophila where regional specification is achieved in a syncitiotblastoderm this process in most other animals must deal with cellular structures. Therefore it is difficult to imagine how the products of the Hox genes which are nuclear proteins could actually specify the positional information in cellular embryos. This last point is important since there are many aspects of murine/mammalian embryogenesis that would preclude the use of some of the mechanisms that have been elucidated in Drosophila.

6.6 Differences between Drosophila and mouse development that could preclude the use of identical genetic mechanisms

In Drosophila the establishment of anteroposterior axis and of the regional specification of points along that axis is achieved through the interaction of a number of transcription factors in the syncitiotblastoderm (see introduction, Akam, 1987, Ingham, 1988). The rostrocaudal axis of the mouse embryo is not established prior to fertilisation but is generated in the
cellular embryo during development, and is first made apparent at
gastrulation. As has been discussed above it is possible that this axis is a
consequence of the orientation of the implanting embryo with respect to its
uterine environment (Smith, 1985). Therefore one would not expect to find the
anteroposterior polarity in the mammal generated by bicoid or oskar like
molecules. The signal(s) for rostrocaudal patterning in cellular embryos is
more likely to be due to molecules that can communicate inter-cellularly. One
would also not expect to find homologous pair rule or segment polarity genes
in the mouse since with the possible exception of rhombomeres there do not
appear to be any segments of developmental significance in the mouse embryo.

That is not to say that genes which are structurally related will not be
found but rather that if these genes are found then they will not have a
homologous function. For example murine genes with homeodomains related to
that of the Drosophila pair rule gene paired have been described although
their expression patterns would seem to be very different than that of the
Drosophila gene (Deutsch et al., 1988). This is also true for the engrailed
like genes (Joyner and Martin, 1987, Davis et al., 1988, Davis and Joyner,
1988, Davidson et al., 1988). A recent study of the expression patterns of the
engrailed related proteins in a number of arthropods, annelids and chordates
suggested that while these proteins probably have related biochemical
functions in these different species they will not have homologous
developmental functions in all the species analysed (Patel et al., 1989).
Although the expression patterns within the arthropods, including Drosophila,
are consistent with a common role in this phylum, in no other phyla were
engrailed related proteins expressed in developing metameres (Patel et al.,
1989).

While it seems that the functions of structurally related proteins in the
mouse and Drosophila may not be functionally conserved it might be possible
that the molecular interactions between some of the proteins will be the same
in both organisms. This might explain why regions of the Dfd, en, and caudal
proteins outside of the homeodomain are conserved. They may help maintain a
useful molecular pathway that could be activated in both Drosophila and mice
but for different reasons.

Another important difference between Drosophila and murine embryogenesis
is that the body plan of the Drosophila embryo is established very rapidly. In
contrast the body plan of the mouse embryo is generated over a larger time
period (see Introduction). With regards to this temporal problem it would be
of interest to examine the events and the molecules involved in the
development of a short germ insect. Short germ band insects, unlike long germ
band insects such as Drosophila, do not have all the segments of the embryo represented on the blastoderm but rather the early embryo only consists of very anterior regions and the more posterior portions are generated over time. In these embryos the regional signals have to cope with a cellular embryo and there is a strong temporal aspect to the regional specification. Therefore in these regards the mechanisms that could be elucidated in these embryos could be more applicable to the mammalian embryo.

6.7 Molecules that could organise Hox gene expression in vertebrates

It is interesting to consider the nature of the hierarchical interactions that occur, and of the molecules that may be involved, in mammalian development and how these signals could result in the observed patterns of Hox gene expression. There are as yet no such candidate molecules in mammalian development but progress has been made in other vertebrate systems, most noticeably in Xenopus. Although Xenopus differs from the amniotes in a number of ways, such as in the induction and organisation of the body plan where it employs more determinative processes and in achieving all the primary body plan specification before growth begins, it is the best understood vertebrate with regards to these processes (Cooke, 1989a). It is also felt that since the body plan is conserved across all early vertebrate development the features described in Xenopus will help clarify the fundamental mechanisms.

6.7.1 Growth factors and regional specification in Xenopus embryogenesis

It has been postulated that there are three signals that are involved in organising the mesoderm of the Xenopus embryo (Dale and Slack, 1987). There are two signals that emanate from the vegetal region, of which one is the ventrovegetal signal which induces the overlying marginal region to become ventrally specified mesoderm and the other is the dorsovegetal signal which induces its overlying dorsolateral marginal zone to become the organiser, which represents dorsal anterior mesoderm. The third signal is that which is thought to be released from the organiser to dorsalize the ventrally specified mesoderm in a graded fashion such that the mesoderm that lies closest to the organiser is most dorsal and that furthest away most ventral. A number of molecules have been identified which can mimic some of these effects. Ventral mesoderm can be induced with fibroblast growth factor, while a TGF-b like factor derived from the XTC cell line, XTC-MIF, can induce dorsal type mesoderm and this tissue can act as an organiser (Smith, 1987, Slack et al., 1987, Cooke, 1989b). The factor for the dorsalisation of the ventral mesoderm may be of a similar biochemical nature to heparin (Dale and Slack, 1987).
Therefore perhaps molecules such as these may activate homeobox gene expression in vertebrate embryos. There have been two recent reports that would indicate that these are the sort of molecules that one should be looking for. Firstly a novel frog homeobox gene, not of the Antp class, was found to be induced in competent ectoderm by mesoderm inducing factor secreted by the XTC cell line (Rosa, 1989). This factor is probably XTC-MIF. This homeobox gene, Mix.1, is not expressed in ectoderm or mesoderm in the embryo but is restricted to endoderm (Rosa, 1989). Secondly there is another Xenopus homeobox gene, Xhox3, which is again not of the Antp class but is most closely related to the Drosophila eve homeodomain, which is predominantly located in posterior regions of the embryo (Ruiz I Altaba and Melton, 1989a, Ruiz I Altaba and Melton, 1989b). This gene has been shown to be down regulated upon XTC-MIF treatment but up-regulated by FGF treatment (Ruiz I Altaba and Melton, 1989b). Thus XTC-MIF which can induce anterior dorsal mesoderm down regulates this gene which is predominantly expressed in posterior regions while FGF which induces more ventral mesoderm acts to increase the level of Xhox3.

These growth factor molecules can clearly regulate the expression of homeobox containing genes, although since these genes are not of the Antp class it is difficult to know how these results relate to the Hox genes. It is possible that genes such as Xhox3 which display broad domains of expression are activated by these growth factor molecules and they in turn activate the genes of the Hox clusters which are differentially expressed along the rostrocaudal axis and generate smaller domains. If this were the situation then this would be very reminiscent of that in Drosophila where the embryo is initially subdivided into large units and then in turn split into even smaller units (see introduction, Akam, 1987, Ingham, 1988).

6.7.2 Retinoic acid and rostrocaudal regional specification

There are other molecules besides growth factors that could be involved in the early stages of regional specification in the vertebrate embryo. It has been recently demonstrated that all-trans retinoic acid has an effect upon the body plan of the developing Xenopus embryo. This molecule has been shown to cause a transformation in the developing central nervous system of Xenopus (Durston et al., 1989). The transformation observed is that of anterior neural tissue to a more posterior specification. One implication from this work is that there exists a posterior source of retinoic acid in the embryo. It is therefore interesting that the chick tail bud mesoderm has ZPA activity when grafted to the appropriate region of the limb bud (Saunders and Gasseling, 1983). This may suggest that there is a localised posterior source
of retinoic acid in the tail bud of the chick and possibly also in mammals.

6.7.3 Retinoic acid and homeobox gene expression

Retinoic acid may be particularly interesting with regards to homeobox genes since many of these genes respond very fast to retinoic acid in teratocarcinoma cells (Colberg-Poley et al., 1985, Hauser et al., 1985, Deschamps et al., 1987, Baron et al., 1987, Murphy et al., 1988). Results from this laboratory (N.Papalopulu pers. comm.) have demonstrated that the Hox 2 genes are very quickly responsive to retinoic acid and that the response is not the same for each member of Hox 2. The speed of the response of these genes to retinoic acid is exceptionally fast, in the order of 20-30 minutes, and therefore these genes may be primary targets for the action of retinoic acid. This fast response is thought to be mediated post-transcriptionally (N.Papalopulu and R.Krumlauf pers. comm.).

If there is a posterior source of retinoic acid in the embryo and the genes are found to be differentially responsive across the cluster one could imagine a scheme that would establish progressively more rostral boundaries of expression for each successive 3’ member of Hox 2. In such a situation one would be predicting that the most caudally expressed gene, such as Hox 2.5, would need the highest concentration of retinoic acid to respond while the more rostrally expressed genes would be able to respond to lower concentrations of retinoic acid and therefore their domains of expression could extend more rostrally away from the source of retinoic acid. Another attractive aspect of this model is that since the homeobox genes encode proteins that are transcriptional regulators these genes once turned on could maintain their own expression through autoactivation. While there is no good evidence in vertebrates for autoactivation of homeobox genes, there is in Drosophila, e.g. Dfd (Kuziora and McGinnis, 1988). Therefore the action of retinoic acid to order the expression of these genes need only be transient since the genes once activated could maintain their own expression. With regards to this it had been noted that when the Xenopus embryos were treated with retinoic acid, and showed a posteriorisation of the front of the nerve cord, the expression of some homeobox genes was upregulated (Durston et al., 1989 and N.Papalopulu pers. comm.).

The relationship between retinol binding proteins and homeobox gene expression is also strikingly demonstrated in the central nervous system. It has been shown that the cytoplasmic retinol binding protein (CRBP), which is expressed in a reciprocal domain to the Hox 2 genes. CRBP is found in the ventral portion of the spinal cord while Hox 2 genes are expressed in a dorsal
domain. These two complementary domains of expression abut. The line which marks the boundary between these two domains does not correspond to any morphological marker. Thus again one can note a relationship between homeobox genes and retinols.

These examples from *Xenopus* serve to point out the sort of molecules that may act as the regional cues in mammalian regional specification. As was discussed in the introduction one approach to the study of development was to utilise molecules that were found to be important in other biological processes, such as cellular transformation, and to analyse their possible developmental roles. Thus it is encouraging that a number of molecules that have been found to be important in regional specification of the *Xenopus* embryo have growth factor properties or are steroid hormones.

6.8 Other roles for homeobox genes

While a role for Hox genes in anteroposterior specification may be that which has been evolutionary conserved and therefore ancient there may also be other roles for these genes in the mouse. The differences in the anterior boundaries in expression between the ectoderm and the mesoderm would suggest that the genes are responding to different signals in the different germ layers. This is substantiated by the fact the the Hox 2.1 gene exhibits a different 5' start site in the lung versus the spinal cord and therefore must be responding to different factors in these different situations (N. Papalopulu pers. comm.). Thus possibly the Hox 2.1 gene, and the other Hox genes that are expressed in the lung, may be performing a function that is related to lung organogenesis. Since these genes are expressed in the mesenchyme, which induces growth and branching on the developing endodermal bud, it is possible that these genes are participating in these functions (Sorokin, 1965). This may also be true in the other organ systems of the mouse.

Another example that may be similar is that of the expression patterns in the transverse plane of the developing nerve cord (Chapter 5). It would seem, that although the rostral boundaries of expression do not change, the transverse pattern of expression is spatially modified as development proceeds. Therefore while there are signals that establish, or have established, the rostral boundary of expression, superimposed upon these restrictions are other signals which cause the observed patterns. These signals may be those that are involved in generating the ordered array of diverse cell types in the nerve cord. If this were so it would be expected that other molecules would display similar restrictions in their expression.
patterns and indeed at least in the case of cytoplasmic retinol binding protein (CRBP) this would appear to be the case (Maden et al., 1989).

It has also been noted that a number of the Hox genes have been found to be expressed in haemopoietic cells (Kongsuwan et al., 1988, Blatt et al., 1988b). The patterns of expression of some of the Hox genes varied in different cell lines (Kongsuwan et al., 1988). Some authors have pointed out altered aspects of homeobox gene organisation or regulation that are found in leukaemic cells. There has been described a deletion on mouse chromosome 2 that is is associated with myeloid leukaemia which spans at least part of the Hox 5 cluster (Blatt and Sachs, 1988a). There has also been reported a genomic rearrangement in a myeloid leukaemic cell line that involves the insertion of an intracisternal A particle upstream of the Hox 2.4 gene (Blatt et al. 1988b, Kongsuwan et al., 1989). This rearrangement results in the constitutive expression of the Hox 2.4 gene in this leukaemic cell line (Blatt et al., 1988b, Kongsuwan et al., 1989).

These observations may indicate that the Hox genes are involved in the differentiation process of haemopoietic cells. Since the Hox gene family represents a large collection of transcriptional regulators that could help define a wide range of diverse cell types it would not be too surprising if during the evolution of the haemopoietic system these genes were recruited.

6.9 Homeobox target genes

It is important to note that while the homeobox genes may be able to self regulate themselves they may also be able to cross regulate each other and modify their patterns of expression. For example if homeobox genes do autoactivate themselves it is possible that other subfamily members, with very similar homeodomains, could compete for the same sites and therefore affect transcription. This could be synergistic or antagonistic. It is clear that, unlike the situation in Drosophila, the posterior murine genes do not repress the expression of the more anterior genes. Thus instead of observing anteriorly and posteriorly restricted domains of expression one observes overlapping domains of expression with different anterior boundaries.

There has been little consideration of what genes homeodomain proteins may activate. Since it is felt that these genes act to interpret positional cues it would seem likely that they would activate molecules that would determine, directly or indirectly, the differences appropriate to a given position. For example, considering the prevertebrae one would not expect the homeobox genes to be implicated in aspects of cellular differentiation that are common to all prevertebrae since different combinations of homeobox genes
are expressed in different prevertebrae. It would seem more likely that the homeobox genes will effect genes which will be involved in the morphogenesis of the tissue.

In the introduction an example of a graft from the hind limb bud to the fore limb bud, in the chick, was cited. In this experiment the grafted material remembered that it came from the hind limb and induced a claw in its new location in the fore limb. It was suggested that this piece of material carried with it a record of its history and that this record could be thought of in terms of a gene network where genes are in an "on" or "off" state and that this record was established after the initial positional signal had been imparted. This is particularly interesting in the light of the observed patterns of homeobox gene expression, not only that of Hox 2 genes but also of the genes of the other clusters. One could think of the Hox genes as being part of the network of genes that can be "on" or "off" which are remembered after the initial positional information has been imparted and are thus part of the record of any tissue. It is certain that the expression of Hox genes in the mouse persists until late stages and even into the adult (e.g. Graham et al., 1988).

6.10 Future experimental approaches to homeobox gene function

While the studies presented in this work support the idea that the murine genes are involved in regional specification they are descriptive and it will be important to couple work of this kind with more embryological and genetic experiments. Hopefully from performing these types of experiments more information could be gained on the roles of Hox genes in murine embryogenesis.

6.10.1 Tissue grafts and Hox gene expression

If these genes are involved in the interpretation of rostrocaudal regional information embryological manipulations could be performed that would confirm this hypothesis. For example, material from the hindbrain region could be transplanted to more caudal portions and the pattern of Hox 2 gene expression examined. This would more specifically mean comparing the expression of Hox 2.5, the caudally expressed gene and that of a more 3′ gene which is more rostrally expressed, such as Hox 2.8.

Since embryonic tissues become gradually restricted in their developmental potency with time, there should be an early period when the transplanted material will respond to its new environment and differentiate accordingly and a later period when it will differentiate as if it were in its original location. These two different situations could be determined by the
use of histological or tissue-specific markers. If Hox 2 genes are involved in the interpretation of rostrocaudal regional specification one would predict that if the graft was done early enough then the piece of hindbrain that was grafted should express Hox 2.8 but not Hox 2.5 but when transplanted to its new location if it responded to this environment this piece of tissue should now switch on Hox 2.5 as a marker of its new posterior location. In the second situation when the transplanted material would not respond to its new environment one would predict that the material after transplantation would not switch on Hox 2.5. In this latter case one would be predicting that the Hox genes are acting as part of the mechanisms that allowed this piece of tissue to remember its previous history.

Given the available methods for mouse embryo culture one could carry out such experiments on the mouse although it would be technically more convenient to work with chick embryos. Since the Hox genes are so strongly conserved (see chapter 3) one would be confident that the results from experiments on the chick embryo would be directly relevant to the mouse embryo.

None of the mouse Hox gene clusters map to any known genetic loci therefore if one is to embark upon a genetic analysis of these genes the mutations must be created. Technology now exists in the mouse to create a range of mutations using DNA micro-injection into one cell embryos or through manipulation of embryonic stem (ES) cells followed by reintroduction into the animal (Rossant and Joyner, 1989).

6.10.2 Generation of gain of function mutants

Micro-injection of Hox gene constructs containing a variety of regulatory sequences into one cell embryos could be used to inappropriately express these genes. In effect a gain of function mutant is being created and one would then screen for a mutant phenotype. This approach has been utilised for two murine homeobox genes Hox 1.4 and Hox 1.1 (Wolgemuth et al., 1989, Balling et al., 1989). The Hox 1.4 gene was introduced into mice and expressed from a construct that contained 10kb of coding and 5' flanking DNA and about 2kb of SV40 sequences. This construct caused the Hox 1.4 gene to be overexpressed and resulted in the development of an abnormal gut although it did not affect other tissue where it was normally expressed (Wolgemuth et al., 1989). The Hox 1.1 gene was put into transgenic mice behind the chicken b-actin promoter (Balling et al., 1989). This ectopic expression of Hox 1.1 resulted in craniofacial abnormalities which were interpreted as being due to the ectopic
expression of the Hox 1.1 gene in the head. The effect of the Hox 1.1 gene is felt to be in the respecification of the positional information received by the cephalic neural crest and therefore results in craniofacial abnormalities since the affected structures are derived from this cell population (Balling et al., 1989).

These results are very encouraging, in particular those with the Hox 1.1 gene, as they strengthen suggestions that the Hox genes play important roles in mammalian development. It will obviously be important to attempt such experiments with the genes of the Hox 2 cluster. One caveat that must be taken into account is that there are suggestions that ubiquitous high level expression of these genes causes embryonic lethality. Balling et al., (1989) have suggested that the mice carry the Hox 1.1 gene behind the chicken β-actin promoter survived to birth because this promoter is so weak.

One possible reason for this supposed lethality is that since the homeobox genes have the capabilities to regulate themselves and to regulate each other the overexpression of any given Hox gene could perturb the whole system by interfering with the normal regulatory network. Therefore overexpression of one Hox gene could result in the repression or activation of a range of other Hox genes and therefore of their target genes. If these genes are involved in the establishment of regional specification then a large scale perturbation of this large gene family and its normal regulation could result in the death of the embryo.

It would be interesting to express any member of Hox 2 throughout the neural tube at all times in development and then to analyse any resultant phenotype. If one could direct expression to a specific region of the embryo, instead of generally overexpressing the gene everywhere, one may avoid lethality or at least delay it until reasonably late periods. Another attraction of this approach is that instead of having to deal with effects of overexpression of a gene in the whole animal, which can be exceptionally difficult to interpret, the overexpression could be restricted to a single organ or tissue where one would hopefully have a better chance of understanding any phenotype that was generated.

This approach could possibly be applied to an analysis of the role to the Hox genes in the development of the central nervous system. This could be achieved by driving expression of any of these genes from any of the neural filament promoters which are active from 9 days of gestation specifically in the neural tube (Cochard and Paulin, 1984). Expression from these regulatory regions should result in general expression of this gene at all places in the developing neural tube and therefore the gene will not exhibit the normal
dynamic patterns of expression in the neural tube. This experiment while expressing the gene at all points across the transverse plane of the neural tube would also result in inappropriate expression of the gene in more rostral regions. Thus one could analyse the effect of inappropriate expression across the transverse plane at positions along the rostrocaudal axis where the gene should normally be expressed and the effects of expressing a gene more rostrally than it normally is.

The roles of homeobox genes in haemopoetic cells could be tested by specifically manipulating cells of this type. One could irradiate an animal to kill its haemopoetic cells and then repopulate this animal with bone marrow cells from another animal that had been infected with a retroviral construct that overexpressed one of the Hox genes. If these genes do have a role in haemopoetic lineages then possibly one could perturb this process. The effect of overexpression of the gene on the differentiation process would be assessed by analysing the cell types and the proportions of each cell type in the blood and tissues of the animal.

Another aspect of these experiments is that one should analyse any transgenic overexpressing a given Hox gene for inappropriate expression of other Hox genes.

One could also attempt to direct the expression of one Hox gene by the regulatory sequences of another. There has as yet been no identification of such regulatory sequences and the regulatory sequences for any gene may be spread throughout the whole cluster (Graham et al., 1989). Experiments in transgenic mice using the $\beta$-galactosidase gene as a reporter gene behind regions from the 5' portion of the Hox 1.3 gene have only shown partial reconstruction of the patterns of expression (Zakany et al., 1988). This would support the above suggestion that the control elements for these genes are spread throughout the cluster. If this is true, to express one homeobox gene in the domain of another the genes would have to be swapped. That is, to express the Hox 2.5 gene in the Hox 2.8 domain then the Hox 2.5 gene would have to be placed in the Hox 2.8 position in the cluster.

One way of achieving this end is to make use of the YAC clone that contains eight members of Hox 2 (see Figure 3.1). By using the recombination systems of yeast, which are very efficient in comparisons to mice, the Hox 2.8 gene, for example, on the YAC could be replaced with the Hox 2.5 gene by performing this recombination in the yeast cells. This recombinant YAC would then be isolated and introduced into mice the phenotype of the resulting embryos analysed. Hopefully this would result in the expression of Hox 2.5 in the normal Hox 2.8 spatial and temporal pattern. Inappropriate expression
could well result in the posteriorisation of regions that normally do not express Hox 2.5. This may result in the alteration of structures in the hindbrain and peripheral nervous system and in the mesodermal organs.

6.10.3 Generation of loss of function mutants

Gain of function mutants can in themselves be confusing and it is important to couple this approach with those that involve the generation of other types of mutation. These experiments hinge upon the use of ES cells. As has been stated in the introduction these cells are derived from the inner cells mass of the embryo, they can be maintained in culture, manipulated and then reintroduced back into the animal and they will contribute to the germ line (Rossant and Joyner, 1989). These cells have been used to create mutations in a number of genes including the Hox 1.1 gene and in the En-2 gene (Zimmer and Gruss, 1989, Joyner et al., 1989). These mutations have been created by transfecting the ES cells with modified versions of the gene of interest and then screening for homologous integration events that have resulted in the replacement of an endogenous copy with the mutated copy. In the case of the Hox 1.1 gene a premature stop codon was introduced into the homeobox and this should result in the production of a truncated protein (Zimmer and Gruss, 1989). Joyner et al., (1989) homologously integrated a neomycin resistance cassette into the En-2 gene and this integration event resulted in the removal of the homeobox. Thus using ES cells one could generate a vast range of different mutations in any gene of interest. In this laboratory and in collaboration with Dr. A.Bradley this technology has been used to disrupt the Hox 2.6 gene (A.Bradley and R.Krumlauf, pers. comm.).

It will be interesting to analyse the embryonic phenotypes once the mutant genes have been passed through the germ line and the mice bred to homozygosity. Mutations that severely disrupt a gene may cause very extreme phenotypes such as early embryonic death and therefore it will be important to create mutations that do not result in such drastic changes in the protein product. Using this system one could create individual point mutations in any gene and generate mice carrying such a mutation. Such animals may have less severe phenotypes that could be very informative.

Having proposed that severe gene disruptions may cause embryonic lethality mice carrying such mutations may well survive until relatively late times. This could happen because the early expression patterns of related subfamily members which share very similar N-termini, hexapeptides and homeodomains, such as Hox 1.4, Hox 2.6 and Hox 5.1, are very similar during gastrulation stages (Gaunt et al., 1989). Thus if one of these genes were
deleted it is possible that the other subfamily members could take over the early functions of this gene, while they are expressed in the same domains. The actual effect of the mutation would then only be felt when the specific patterns of expression of the mutated gene arose.

There is precedence for this sort of situation in *Drosophila*. The *Ubx* and *abd-A* genes of *Drosophila* are the most closely related genes of the homeotic genes in their homeodomains (Akam et al., 1988). These two genes exhibit some redundancy where their expression patterns overlap, such as in the nervous system, and in these cases the activity of either gene will suffice (Ghysen and Lewis, 1986). There has also been described a mutation in the BX-C which results in a deletion of part of the *Ubx* and *abd-A* genes (Rowe and Akam, 1988, Casanova et al., 1988). This mutation actually causes a fusion protein to be produced which consists of the amino terminal part of *abd-A* and the carboxy terminal of *Ubx*. Embryos of this type, which lack *abd-A* and *Ubx* genes but carry the fusion *abd-A/Ubx* gene can develop relatively normally with this fusion protein being able to rescue many of the functions of either of these two proteins. Thus proteins with similar but not identical homeodomains can compensate for other homeodomain proteins that are absent.

These embryological and genetic approaches should begin to test many of the ideas presented in this thesis. Besides providing information on homeobox gene they will also be invaluable to any studies on mammalian embryology. These experiments and those by other labs, on other interesting molecules, will hopefully be able to begin to address the hierarchy of interactions that are involved in regional specification in mammalian development.
REFERENCES


Dolle, P., and Duboule, D. (1989). Two gene members of the murine Hox-5 complex show regional and cell-type specific expression in developing limbs and gonads. EMBO J. 8: 1507-1515


Rugh, R. (1968). The mouse: its reproduction and development. Burgess


Schughart, K., Pravtcheva, D., Newan, M.S., Hunihan, L.W., Jiang, Z., and Ruddle, F.H. (1989). Isolation and regional location of the murine homeobox-containing gene Hox 3.3 to mouse chromosome region 15E. Genomics 5:


132


The Murine and Drosophila Homeobox Gene Complexes Have Common Features of Organization and Expression

Anthony Graham, Nancy Papalopulu, and Robb Krumlauf
Division of Eucaryotic Molecular Genetics
National Institute for Medical Research
The Ridgeway, Mill Hill
London NW7 1AA
England

Summary

In situ hybridization analysis of mouse embryos shows the seven members of the Hox-2 complex to be differentially expressed in the central and peripheral nervous system and in mesodermal derivatives (somites and lung). Beginning at the 5' end of the cluster, each successive gene displays a more anterior boundary of expression along the central nervous system. A gene's position in the Hox-2 cluster therefore reflects its relative domain of expression along the anteroposterior axis of the embryo, a feature observed with Drosophila homeotic genes. Sequence comparisons of the Hox-2 cluster with other mouse and Drosophila homeobox genes have defined subgroups of related genes; in the mouse there are four clusters related by duplication and divergence. Alignment shows a clear relationship among genes in the mouse and Drosophila complexes, based on relative position, sequence identity, and domains of expression along the rostral-caudal axis. Our results argue that these complexes arose from a common ancestor, present before the divergence of lineages that gave rise to arthropods and vertebrates.

Introduction

Drosophila genes implicated in specifying segmental pattern and segment identity, such as engrailed and Antennapedia (Antp), have been found to contain a common element, the homebox (reviewed in Gehring, 1987). An initial study of the phylogenetic distribution of Antp-related homeboxes suggested that the presence of this sequence in the genome of an animal may correlate with a segmented body plan (McGinnis, 1985). Yet it was subsequently shown that homeoboxes are evolutionarily more widespread, with copies of this element being found in the genomes of animals from widely separate phyla including nonsegmented invertebrates and mammals (Holland and Hogan, 1986). Of the genes in which the homebox was first discovered, only fushi tarazu is a segmentation gene, while the others are homeotic genes believed to be responsible for anteroposterior specification (Slack, 1984).

Later studies uncovered Drosophila homebox-containing genes that exhibit patterns of expression that suggest roles in positional information independent of the process of segmentation (Hoey et al., 1986). In the case of Antp-related homeoboxes, this role may be in the specification of positional values along the anteroposterior axis of the animal. Thus, it would seem that homeoboxes are involved not only in the process of segmentation but also in specifying positional information.

Support for a role of homeobox genes in mouse development is derived from their patterns of expression observed in the embryo. Homeobox genes are expressed in derivatives of the ectoderm (both central nervous system [CNS] and peripheral nervous system [PNS]) and mesoderm (lung, somites, gut, and kidney), but generally not in endoderm (Feinberg et al., 1987; Krumlauf et al., 1987; Dressler and Gruss, 1988; Graham et al., 1988). The genes also display similar but not identical overlapping domains of expression within the same tissue (Gaunt et al., 1986, 1988; Graham et al., 1988; Holland and Hogan, 1988), which suggests that they may have a role in the establishment of positional information.

Many of the Drosophila homebox-containing genes have been found to be clustered in two separate but adjacent complexes, Antennapedia (ANT-C) and Bithorax (BX-C), which are located on the right arm of chromosome three (Lewis 1978; Kaufman et al., 1980; Sánchez-Herrero et al., 1985). One important feature of these complexes is that the physical order of the homeotic genes in the clusters is identical to the order in which these genes are expressed along the anteroposterior axis of the embryo during development (Harding et al., 1985; Akam, 1987). The domains of expression along this axis also correspond directly to the segments that are affected by mutations in these genes (Harding et al., 1985). Therefore, there is a correlation in Drosophila between the relative position of a gene within a cluster, its domain of expression, and its effects on the specification of structures along the anteroposterior axis of the embryo (Harding et al., 1985; Akam, 1987).

The organization of homebox genes in clusters is a general phenomenon that has also been described in mouse (Duboule et al., 1986; Hart et al., 1987b; Graham et al., 1988; Breier et al., 1988; Sharpe et al., 1988), human (Boncinelli et al., 1988), Xenopus (Harvey et al., 1986), and zebra fish (Njolstad et al., 1988). The murine genome contains more Antp-related homeoboxes than Drosophila, with at least 25 members identified to date in this large gene family. These genes are organized into four clusters, termed Hox-1, Hox-2, Hox-3, and Hox-5, located on chromosomes 6, 11, 15, and 2, respectively (Bucan et al., 1986; Hart et al., 1985; Breier et al., 1988; Featherstone et al., 1988). Analysis of the sequence and organization of genes from these clusters revealed that some members had substantial identity in several regions of their predicted proteins in addition to the homeodomain (Regulski et al., 1987; Featherstone et al., 1988; Graham et al., 1988). These highly related members form a subgroup, and based on their identity and alignment within the clusters, it has been suggested that the mouse clusters are related to each other by the duplication and divergence of an ancestral cluster (Hart et al., 1987b;
The high degree of conservation in both sequence and organization of the mouse homeobox gene clusters is intriguing, and may by analogy to the Drosophila clusters reflect aspects of their regulation or function. Thus, it is important to examine the existence of such relationships in the same and different organisms. Toward this end, we have analyzed members of the Hox-2 complex with respect to correlations between their sequence, organization, and patterns of expression in the mouse embryo. Such a correlation has been found to exist and has revealed important similarities between the mouse and Drosophila homeobox gene complexes. These observations shed light on the possible evolutionary history of homeobox gene clusters. We suggest that both the mouse and Drosophila homeobox clusters have arisen from a common, ancient ancestral complex. This ancient cluster of genes is related to both its murine and Drosophila counterparts by the sequences and patterns of expression of the genes along the anteroposterior axis of the animal.

Results

Organization and Expression of Hox-2
We have recently reported the isolation of cosmid clones spanning the Hox-2 complex on chromosome 11, which contains seven homeobox genes (Graham et al., 1988). These clones have now been used to prepare subcloned genomic fragments for each gene and to isolate cDNAs, which have allowed us to generate gene-specific probes for analyzing the structure, sequence, and expression patterns of the Hox-2 complex. Figure 1 shows the ordering of the genes in the Hox-2 complex (the numbering reflects the order in which the genes were discovered). All seven genes have the same 5' to 3' orientation with respect to transcription, as revealed by sequence and expression data, with Hox-2.5 as the most 5' member.

A comparison of the expression in embryonic tissue for adjacent genes in the Hox-2 cluster is shown in Figure 1. Because of the high degree of identity between many mouse homeobox genes, we used probes for each gene that would minimize cross-hybridization. Clearly, every member of the cluster is expressed during embryogenesis, and in RNA from 12.5 day embryos many of the genes display complex patterns of transcription as evidenced by multiple transcripts that differ in both size and intensity. There are tissue-specific differences in the transcript distribution of some genes, as seen for Hox-2.6 in Figure 1. The two major transcripts, of 2.4 and 4.1 kb, are observed in the RNA from whole embryos; however, the 4.1 kb band is reduced in relative intensity in the spinal cord and is barely detectable in lung. We do not know to what extent the multiple bands represent differential splicing or multiple start sites, and it is possible that these transcripts...
could encode more than one protein. However, in the case of Hox-2.1 we have found that there are multiple initiation sites that differ between tissues (Garbern et al., submitted).

Most mouse homeobox genes are expressed in ectodermal derivatives (CNS and PNS), and all members of Hox-2 show high levels of expression in RNA from 14.5 day embryonic spinal cord (Figure 1). The expression in the spinal cord of all Hox-2 genes persists throughout development up to birth, and, in the one case we have examined, expression in the spinal cord continues in adult stages (Krumlauf et al., 1987). Expression of the complex in a mesodermal derivative, lung, shows a different pattern. The three most 5' members of the cluster, Hox-2.3-2.5, are not expressed in this tissue, while the four genes at the 3' end of the complex display high levels of expression in lung. The lung is generated from anteriorly derived mesoderm, and these results support the idea (see Discussion) that expression in anterior structures correlates with genes located in the 3' part of the Hox-2 cluster.

In Situ Analysis of Hox-2 Gene Expression along the Anteroposterior Axis

The Hox-2 complex is expressed (as shown above) in the CNS, which is an ectodermal derivative that lies along the length of the anteroposterior axis of the animal. We have used in situ hybridization to examine the spatial restrictions in the patterns of expression of every member of Hox-2 within the CNS along the anteroposterior axis. Figure 2 shows an example of these results with the Hox-2.2 gene in a sagittal section of a 12.5 day mouse embryo. There is an anterior boundary in the domain of expression in the CNS that maps to the hindbrain (sc). However, in this section and other serial sections there is no clear posterior limit to this expression. A similar type of spatial restriction is also observed in the somitic mesoderm, where expression extends from the most posterior prever-
tebrae to an anterior boundary corresponding to the seventh prevertebra (Figure 2). The anterior cutoff of expression for \textit{Hox-2.2} within the somitic mesoderm is more posterior than in the CNS, and these overlapping but offset domains of expression in ectoderm versus mesoderm appear to be a common feature of homeobox expression in the mouse (Gaunt, 1988; Gaunt et al., 1988; Dony and Gruss, 1987).

Using all the \textit{Hox-2} probes, we have observed overlapping but distinct domains of expression for each gene in the CNS. In view of the absence of a clear posterior boundary of expression, we have focused on examining the different anterior boundaries of expression for each member of the \textit{Hox-2} complex. We consistently observe the same relative anterior boundaries of expression in all 12.5 day mouse embryos. All seven genes have been directly compared in serial sections, using relative distances from morphological structures to map the boundaries. Starting with \textit{Hox-2.5}, we find that as one progresses through the \textit{Hox-2} complex in a 5' to 3' direction, the limit of expression of each gene in the CNS becomes progressively more anterior. Figure 3 shows that for the most 5' gene, \textit{Hox-2.5}, the limit of expression maps within the spinal cord at the level of the third prevertebra. The boundary of the adjacent \textit{Hox-2.4} gene is within the posterior myelencephalon (hindbrain), as are those of the more 3' genes \textit{Hox-2.3} (not shown) and \textit{Hox-2.2} (Figure 3). The differences in the anterior boundary between \textit{Hox-2.5} and \textit{Hox-2.4} are large by comparison with the differences between \textit{Hox-2.2} and \textit{Hox-2.4}; however, each of the genes does have a successively more anterior limit (see Figure 6).

Figure 4 illustrates, in serial sections of an embryo, that this general trend continues throughout the complex. Domains of expression for members 3' of \textit{Hox-2.2} extend more rostrally within the myelencephalon, and there is a relatively large difference between the anterior boundaries of \textit{Hox-2.2} and \textit{Hox-2.1}. The successively more rostral expression in the hindbrain for \textit{Hox-2.1-2.7} is shown in Figures 4 and 5. A summary of these in situ results is illustrated in Figure 6. Based on these findings we conclude that the position of a gene in the \textit{Hox-2} cluster reflects its relative domain of expression along the anteroposterior axis of the embryo in the CNS. In similar in situ hybridization experiments, this pattern appears to be already established in earlier embryos (9.5 day). This correlation between position and expression in \textit{Hox-2} also appears to apply to patterns of expression in somitic mesoderm (prevertebrae) and is consistent with expression in lung (Figure 1).

This correlation in limits of expression is analogous to the Drosophila BX-C and ANT-C homeobox gene clusters, where the physical order of the genes along the chromo-
some is related to the anteroposterior order of the segment that recessive mutations of each gene affect (Harding et al., 1985; Akam, 1987). A similar genetic correlation has also been described for the homeobox gene cluster HOM-C of Tribolium castaneum (Beeman, 1987). In situ hybridization analysis of the expression of the Drosophila homeotic genes in BX-C and ANT-C also reveals a relationship between the transcript distribution of each gene in the CNS and its relative chromosomal position (Harding et al., 1985; Martinez-Arias et al., 1987; Mlodzik et al., 1988). These studies have shown that the anterior boundary of Abdominal-B (Abd-B) expression maps to the sixth abdominal ganglion, while Abd-A transcripts extend to the first abdominal ganglion and Ultrabithorax (Ubx) expression reaches the metathoracic ganglion. The homeotic genes of ANT-C maintain this trend with Antp showing expression to the prothoracic ganglion, and Sex combs reduced (Scr) to the second subesophageal ganglion. Deformed (Dfd) transcripts can be detected in the first subesophageal ganglion (Harding et al., 1985), and labial (lab) has been reported to be expressed more anteriorly than Dfd (Mlodzik et al., 1988).

**Homeodomain Sequence Comparison between Murine and Drosophila Clusters**

Recently we reported (Graham et al., 1988) that the Hox-2.6 gene shares substantial sequence identity in multiple regions of the protein with other mouse (Featherstone et al., 1988) and vertebrate homeobox genes (Harvey et al., 1986). Furthermore, Hox-2.6 formed part of a group of mouse genes that displayed significant identity with the Drosophila Dfd gene (Regulski et al., 1987). Based on these identities and the similarities in the expression pattern described above, we wanted to examine whether other genes in the Hox-2 cluster might also show relatedness to Drosophila homeobox genes (Krumlauf et al., 1987; Graham et al., 1988). Sequences of the Hox-2.2 and Hox-2.3 genes have been published (Hart et al., 1987b; Meijlink et al., 1987; Schughart et al., 1988), and we have now isolated and sequenced all the homeoboxes and most of the predicted proteins of the remaining members of the Hox-2 complex (Krumlauf et al., 1987; Graham et al., 1988; Figure 7; unpublished data).

In Figure 7 we compare the sequences of all the Hox-2 homeodomains and their closest murine and Drosophila
Hox-2.5, -1.7, -3.2, (ANT-C, BX-C) counterparts against Drosophila homeodomain. This comparison clearly shows changes and positions of the changes (when compared to Drosophila homeodomains, it is also clear that many of these sequences show similarity to one subfamily. The comparisons between Abd-B and Hox-2.5, Scr and Hox-2.1, Dfd and Hox-2.6, and lab and Hox-1.6 show the highest degree of sequence identity. The Hox-2.7 subfamily shows identity with the zerknüllt (zen) homeodomain, but zen is not a homeotic gene; this group may be more related to proboscipedia (pb; Pultz et al., 1988), which has a homeodomain very similar to that of zen (Cribbs and Kaufman, personal communication). The remaining Drosophila genes including Abd-A, Ubx, and Antp are all so similar in sequence that it is difficult to assign these genes to any of the remaining murine subfamilies, which are represented by Hox-2.4, -2.3, and -2.2. Hox-2.6 displays several regions of identity, outside the homeodomain, when compared with the Dfd protein (Graham et al., 1988), and Mlodzik et al. (1988) have found extended regions of homology between lab and Hox-1.6 (Duboule et al., 1986; LaRosa and Gudas, 1988). The Scr protein (LeMotte et al., 1989) shares identity with Hox-2.1, particularly in the amino-terminal region, where 10 of the first 13 amino acids are the same. Therefore, where protein sequence is available, mouse and Drosophila genes identified as subfamily members, based on similarities within the homeodomain, also display some identity in other domains.

One striking feature of these comparisons is that if we align the mouse and Drosophila subfamilies, then the physical order of the Drosophila homeobox genes along the chromosome and that of their closest related murine genes are identical (Figure 7). These results suggest that the murine and Drosophila homeobox gene clusters are derived from a common ancestor that arose before the divergence of the lineages that gave rise to arthropods and vertebrates. Akam et al. (1988) have recently proposed that the earliest ancestor for the myriapod-insect lineage possessed a homeobox cluster with distinct genes related to lab, pb, Dfd, Antp, and Abd-B. Subsequently, by duplication and divergence the Antp gene gave rise to the Scr gene and then, more recently, to the Ubx and the Abd-A genes. In our comparisons the best overall identities between mouse and Drosophila are in those genes (lab, Dfd, Antp, and Abd-B) predicted to be the most ancient by Akam et al. (1988). In addition, the identity between Scr and Hox-2.1 argues that Scr was a part of the ancestral cluster. It appears that both flies and mice have four genes related to Antp. These could be derived, wholly or in part, from duplication events that occurred prior to the separation of the lineages that gave rise to arthropods and vertebrates, and have been part of the ancestral cluster. They may also have arisen by independent events in each of the major lineages, as has been postulated for the engrailed genes (Dolecki and Humphreys, 1988; Joyner and Martin, 1987). Even if the Antp group of genes arose independently, our results argue for a common evolutionary origin between the mouse and Drosophila clusters.

Discussion

The Homeobox Gene Complex

In this report we have compared aspects of the structure and expression of the Hox-2 complex with other mouse...
and Drosophila homeobox gene clusters. Based on identity shared within homeobox sequences and their predicted proteins, it is possible to divide the murine genes into subfamilies, also related by specific common differences they share when compared with the \( \text{Antp} \) homeodomain (Figure 7). We and others have used this type of comparison, in part, to support the idea that the mouse \( \text{Hox-2} \) and \( \text{Hox-1} \) clusters are related by duplication events (Fibi et al., 1988; Graham et al., 1988; Hart et al., 1987a; Papalopulu et al., 1989).

In light of the newly defined \( \text{Hox-5} \) cluster (Duboule and Dolle, 1989) and the assignment of \( \text{Hox-6.1} \) and -6.2 to the \( \text{Hox-3} \) cluster on chromosome 15 (Schughart et al., 1989), it appears that duplication and divergence of homeobox genes have lead to at least four evolutionarily related complexes in the mouse. In each of the four clusters, the relative position of each subfamily member (Figure 7) and the physical order and spacing of the genes are very similar. This allows us to align the \( \text{Hox} \) clusters into an evolutionarily related complex, as shown in Figure 8. The different members of a related subfamily are represented by the vertical rows of solid boxes. For example, \( \text{Hox-2.5}, -1.7, -3.2, \) and -5.2 form one such family related to the Drosophila \( \text{Abd-B} \) gene. The alignment also reveals that not all genes or subfamilies are represented in each mouse cluster, and suggests that either some clusters arose by partial duplication or there has been sufficient divergence to make some genes undetectable. Using this alignment we might also predict the potential position and existence of other genes in the parts of the clusters that have not been cloned. Based on the extensive similarity between the \( \text{Hox-2} \) and \( \text{Hox-1} \) clusters, we isolated clones 3' of \( \text{Hox-2.7} \) while looking for a \( \text{Hox-1.6} \)-related gene, and identified a new gene (as yet unsequenced) that we term \( \text{Hox-2.8} \). In the same manner we have looked for genes 5' of \( \text{Hox-2.5} \) that would be related to \( \text{Hox-5.3} \). We have walked 30 kb 5' of \( \text{Hox-2.5} \), but have no evidence for additional genes.

**A Common Ancestor for Mouse and Drosophila Homeobox Complexes**

A remarkable feature of our comparisons with mouse and Drosophila homeodomains is the substantial similarity between many of the mouse subfamilies and genes in the ANT-C and BX-C complexes (Figure 7). The identity is based on common conserved changes within the homeodomains; however, in several cases where sequence data are available these comparisons reveal that mouse and Drosophila members in the same subfamily (\( \text{Dfd}, \text{Hox-2.6}; \text{Scr}, \text{Hox-2.1}; \text{lab}, \text{Hox-1.6} \)) also have conserved regions in other domains of the predicted proteins (Graham et al., 1988; Featherstone et al., 1988; Mlodzik et al., 1988; LeMotte et al., 1989). There is a strong correlation between both the position of genes and their relatedness.
within the mouse and Drosophila homebox clusters (Figure 7). In Figure 8 we have positioned the Drosophila homeobox genes above the four mouse clusters and indicate the most closely related subfamilies to illustrate this point. It is difficult to make an absolute one-to-one correlation between the mouse and Drosophila genes, but subgroups of Abd-B, Scr, Dfd, pb, and lab are observed. However, because the three Drosophila genes (Antp, Ubx, and Abd-A) are all highly similar to each other in the homeodomain, direct alignment of these with the mouse genes Hox-2.2, -2.3, and -2.4 is uncertain.

It is important to note that Akam et al. (1988) have proposed that the ancestor of the Drosophila homeobox cluster possessed a set of four or five distinct genes of the lab, pb, Dfd, Antp, and Abd-B type, and that the Ubx and Abd-A genes arose by more recent duplication from the Antp gene. The Scr homeobox is also very similar to Antp, and it is not clear if Scr resulted from an early duplication separate from events leading to Abd-A and Ubx or was part of the primordial cluster. The best sequence identities between the mouse and Drosophila genes in Figures 7 and 8 are observed for the set of genes predicted to be members of the ancestral cluster of Akam et al. (1988). In the mouse there also appear to be four genes related to Antp (Hox-2.1–2.4). This may suggest that these genes arose by independent duplication of the ancestral Antp-type gene also in the lineage that yielded chordates or that they are derived from a similar, earlier duplication event in a common ancestor. This earlier event would have produced more than one Antp-type gene in the ancestral cluster prior to the divergence between arthropods and vertebrates. We favor the latter possibility at least in the case of Hox-2.1, which also shows identity with Scr and supports the idea that Scr was a part of this common ancestral cluster. Interestingly, this group of Antp-related genes in the mouse also appears to be the most variable part of some clusters, as illustrated in the Hox-5 cluster where there are no members of the Abd-A, Ubx, Antp, or Scr subfamily (Duboule and Dolle, 1989).

The sequence identity between the mouse and Drosophila genes does not necessarily imply that these mouse genes are true vertebrate homologs of the Drosophila homeotic genes. Some subfamilies have four members related to a given Drosophila gene. As such, a precise murine equivalent would be hard to establish. These comparisons suggest that the murine and Drosophila gene clusters arose from an ancestral cluster that existed prior to the divergence of the lineages that eventually gave rise to mice and flies. This is supported by the observation that some regions of the proteins besides the homeodomain have been conserved (Regulski et al., 1987; Graham et al., 1988). One difficulty with this argument is that the Drosophila genes are divided into two separate clusters although both are on chromosome three. However, in a more primitive insect, the red flour beetle Tribolium castaneum, the ANT-C and BX-C equivalents are found in a single homeotic cluster, HOM-C (Beeman, 1987). Therefore, ANT-C and BX-C may have originated from one cluster separated by a more recent evolutionary event occurring in certain insects.

Positions of Genes in a Cluster Reflect Domains of Expression in the Rostral–Caudal Axis for Both Mouse and Drosophila

In this paper we have demonstrated that all of the members of the Hox-2 complex are expressed in slightly different overlapping domains in the spinal cord, with respect to the anteroposterior axis (Figures 2–6). We have not detected any posterior limits for expression of these genes in the CNS. However, as one moves from the most 5’ member of the cluster (Hox-2.5) in a 3’ direction, each successive gene displays a more anterior boundary of expression (summarized in Figure 6). The cutoff for Hox-2.5 in the CNS maps opposite the third prevertebra, and the other six genes have anterior boundaries in the myelencephalon. Our findings clearly show that the position of a gene in the Hox-2 cluster is correlated with its relative domain of expression in the anteroposterior axis of the embryo, suggesting that these homeobox genes may be playing a role in the establishment of positional information. This is analogous to the pattern observed for the Drosophila BX-C and ANT-C clusters. The physical order of the genes in these clusters reflects the anteroposterior order of the segments each gene affects and their relative domains of expression.

Our results are in agreement with the recent findings of Gaunt et al. (1988) and Duboule and Dolle (1989) showing a similar correlation between the position of some genes within other murine clusters and their relative extent of expression along the rostral–caudal axis. In all four of the murine homeobox clusters, the genes located at the 5’ end are expressed in more posterior domains than genes located in the 3’ part of the same cluster. It will be important to examine to what extent murine members of the same subfamily all share similar anteroposterior domains of expression. The Dfd subgroup (Hox-2.6, -1.4, -5.1) appears to have very similar patterns of expression (Gaunt, Duboule, and Krumlauf, unpublished data). However, in the Abd-B family Hox-2.5 is expressed in more anterior regions of the CNS than Hox-5.2 (Duboule and Dolle, 1989), and both Hox-2.5 and -2.4 are expressed more anteriorly than Hox-3.1 (Breier et al., 1988). This suggests that genes in the different clusters that are members of the same subfamily do not necessarily have identical domains of expression.

An important question from our results concerns why many of the Hox-2 genes have anterior boundaries of expression in the myelencephalon. At 12.5 days these boundaries do not appear to correspond to any morphological structures within the hindbrain. However, in early stages of neural development, periodic swellings along the axis of the neural epithelium have been observed, and these structures are termed neuromeres (rhombomeres in the hindbrain). Previously, it had not been clear whether segmentation of the neural tube, as defined by the rhombomeres, plays any role in establishing the underlying pattern of development in the hindbrain. Recently, Lumsden and Keynes (1989) have presented data suggesting that rhombomeres do play a significant role in the segmental patterns of neuronal development in the chick hindbrain. Consistent with these results, Wilkinson et al. (1989) have shown that rhombomeres represent domains of expres-
sion for the zinc finger gene Krox-20, providing molecular support for segmentation of the CNS.

The repeated array of spinal motor and sensory roots and sympathetic ganglia in the PNS is a consequence of segmentation in the somitic mesoderm and is not an intrinsic property of the spinal cord or the PNS (Keynes and Stern, 1988). In the hindbrain the alternating pattern of cranial nerve nuclei and their association with particular rhombomeres (Lumsden and Keynes, 1989) is established in the absence of input from somitic mesoderm, implying that segmentation in the hindbrain proceeds by a different mechanism. Therefore a mechanism involving inherent segmentation of the neural epithelium may be significant, particularly in the hindbrain which serves as a complex coordination center, and help to establish the regional differences between the hindbrain and the spinal cord. The expression of Krox-20 and many of the homeobox genes may be involved in achieving this regional specification. The different boundaries of Hox-2 genes could represent limits of expression that correspond to different rhombomeres in the hindbrain, suggesting that they may play a role in establishing the rhombomeric pattern. It will be important to make direct comparisons at early stages of mouse development with Krox-20 to determine whether these anterior limits represent rhombomere boundaries.

**Similarities of Mouse, Other Vertebrate, and Drosophila Complexes**

In both mouse and Drosophila, anteroposterior expression of homeobox genes in the CNS correlates with the organization of the complexes. But what is even more striking is that the homeobox sequence of the most posteriorly expressed member of Hox-2 (Hox-2.5) is most closely related to the most posteriorly expressed member of BX-C (Abd-B). This correlation holds for the other genes in both the mouse and Drosophila clusters. From this we conclude that there is a clear relationship between relative position, sequence identity, and domains of expression along the anteroposterior axis for both the mouse and Drosophila homeobox gene networks, which strongly argues for a common ancestral origin of these complexes.

These similarities with Drosophila are not unique for the mouse. Homeobox genes have been found to be organized in clusters in all vertebrates so far analyzed (see Introduction), and the overall structure is very similar to that of the mouse clusters. Boncinelli et al. (1988) have shown that the human homeobox complexes are organized as in mouse and display identity to Drosophila. The chicken genome has Hox-2 and Hox-1 complexes that are highly related to those of the mouse in both structure and expression (Fainsod and Kuroiwa, personal communication; Wedden et al., 1989). The high degree of identity between the chick and mouse Hox-2.1 and Hox-2.6 genes allowed us to examine their expression by Northern blot analysis (Papalopulu et al., 1989) and in situ hybridization in chick embryos (A. G., unpublished data). Our preliminary results suggest that the patterns of expression in chick embryos are nearly identical. In zebra fish a Hox-2 cluster has also been isolated. One member, ZF-21, displays 85% homology at the amino acid level over the entire protein with Hox-2.1 and is expressed in anterior parts of the CNS (Njolstad et al., 1988; personal communication). It has been shown that related subfamilies are also present in many vertebrates, including Xenopus (Regulski et al., 1987; Graham et al., 1988; Harvey et al., 1986). Fritz et al. (1989) have linked three Xenopus genes (Xhbox 6, 7, and 2) and suggest that they represent the Xenopus homologs of three mouse Hox-2 genes (Hox-2.5, -2.4, and -2.3). The Xhbox 2 and mouse Hox-2.3 proteins have 75% homology over the entire protein. Therefore we suggest that homeobox complexes in all vertebrates may have a structural and functional organization analogous to that we have described for the mouse.

How can we account for the strong correlation between genetic linkage and domains of expression in the mouse and Drosophila homeobox clusters? The conservation in expression and organization we believe to be a consequence of cis-acting regulatory elements distributed along a cluster. If the clusters are broken apart, they would then lose their appropriate coordinated expression. We suggest that each gene may be controlled by multiple elements over a distance, and the normal pattern of gene expression results from the additive effects of these elements. Consistent with this idea is the fact that complex transcription patterns involving multiple transcripts and initiation sites are observed for many of the Hox-2 genes, and in the human Hox-3 cluster differential splicing of at least three genes is observed (Simeone et al., 1988). Zakany et al. (1988) have also found that transgenic mice carrying Hox-1.3 constructs are only able to generate part of the normal spatial pattern of regulation in the nervous system and do not express in mesodermal tissues. These constructs contained large amounts of flanking DNA, implying that other sequences in the cluster are required for normal regulation. Therefore, we feel that the maintenance of homeobox gene clustering in vertebrates is a result of selection based on regulatory mechanisms. In this way each of the mouse clusters may have a slightly different type of regulation, and closely related proteins with potentially similar functions may be independently regulated in some tissues. This may represent a mechanism, different from the evolution of new genes, whereby existing genes are recruited for new purposes by altering control processes.

**Experimental Procedures**

**Hybridization Probes**

All probes from the Hox-2 cluster were subcloned into the Bluescript vector for double-strand sequencing using Sequenase (United States Biochemical Corp.) according to the suppliers and for in vitro transcription with either T7 or T3 RNA polymerase. For in situ hybridization the probes were made with 32P-UTP and for blotting analysis with [32P]UTP supplied by New England Nuclear (Du Pont). Probes were as follows: Hox-2.5 (a 3'400 bp Bgill–Dral fragment), Hox-2.4 (1 kb SacI fragment), Hox-2.3 (800 bp EcoRl–BamHI fragment), Hox-2.2 (1.2 kb EcoRl–SacI fragment), Hox-2.1 (800 bp SacI–EcoRl fragment), Hox-2.6 (800 bp SalI–Bgill fragment), and Hox-2.7 (1.4 kb SacI–BamHI fragment). The Hox-2.1 and Hox-2.6 probes were those reported by Krumlauf et al. (1987) and Graham et al. (1988).

**RNA Isolation and Northern Hybridization**

Poly(A)+ mRNA was isolated from mouse tissues essentially as described in Krumlauf et al. (1987). In brief, the samples were homo-
tion, redissolved in 10 mM Tris-HCl (pH 7.6), 1 mM EDTA, 0.5% SDS, and extracted several times with phenol-chloroform. Following ethanol precipitation, the poly(A)^* mRNA was isolated by oligo(dT)-cellulose chromatography. The RNA samples were electrophoresed in denaturing formaldehyde–agarose gels, transferred to GeneScreen (Du Pont) in 20% SSC, and coupled to the membrane by ultraviolet cross-linking and baking as described in Krumlauf et al. (1987). The filters were hybridized using cRNA-labeled single-stranded antisense RNA probes produced by in vitro transcription of subconfluent gene-specific fragments with T7 or T3 RNA polymerase. Hybridization was in 50% formamide, 5x SSC, 0.1% bovine serum albumin, 0.1% Ficoll 400, 0.1% polyvinylpyrrolidone, 20 mM sodium phosphate (pH 6.8), 1% SDS, 7% dextran sulfate, 100 µg/ml tRNA, 10 µg/ml poly(A) at 65°C for 16 hr. The filters were washed in 0.1x SSC, 0.5% SDS at 80°C for 2 hr, and in the cases where high-stringency conditions were required to ensure that the signal was not derived from related genes, the filter was further treated with RNAase A. The filters were incubated in 2x SSC, 0.1% SDS at 52°C for 1 hr.

In Situ Hybridization

The protocol used was basically that of Wilkinson et al. (1987) with some modifications, and is essentially as follows. Mouse embryos were fixed in 4% paraformaldehyde in PBS overnight at 4°C, and then embedded in paraffin wax. Sections (6 µm) were cut and dried onto gelatin-subbed slides. The section were dehydrated in xylene, treated with 0.2 M HCl, rehydrated in paraformaldehyde, subjected to proteinase K treatment, and then treated with acetic anhydride. After dehydation, probes were redissolved at a final activity of 1–2 x 10^5 dpm/µl in hybridization buffer (50% formamide, 0.3 M NaCl, 20 mM Tris, 5 mM EDTA, 10 mM sodium phosphate, 10% dextran, 1x Denhardt’s solution, 0.5 mg/ml tRNA). Hybridization was overnight at 55°C. Sections were washed at 65°C in 50% formamide, 2x SSC, 100 mM EDTA for 40 min followed by incubation with RNAase A at 20 µg/ml in 0.5 M NaCl, 10 mM Tris, 5 mM EDTA for 30 min. Sections were again washed in 50% formamide, 2x SSC, 100 mM EDTA at 55°C for 30 min followed by two 15 min washes in 2x SSC and 0.1x SSC. Sections were dehydred through alcohol solutions containing 0.3 M ammonium acetate. Slides were dipped in a mix of ilford K5 nuclear emulsion and glycerox–water (6 ml emulsion in 8.82 ml H2O and 0.18 ml glycerol) and kept at 4°C until developed. Sections were stained in 0.02% toluidine blue for 1 min and then mounted in Permount.

Sequences and Organizational Data Used for Comparisons

The references for the four mouse clusters and the Drosophila homeobox genes: some aspects of their organization and structural and functional organization resembles that of Drosophila homeotic genes. EMBO J., in press.


References


Acknowledgments

We thank Paul Hunt and Jo Lorimer for homoeodomain sequences. We also thank Denis Dubreoule for the generous sharing of ideas and data prior to publication, and Steve Gaunt, Peter Gruss, Dado Boncinelli, Paul Sharpe, and Abraham Fainsod for valuable discussions. We are grateful to Leanne Wiedemann, David Wilkinson, and Andrew Lumsden for discussions on data and critical reading and comments on the manuscript. N. P. was supported by the Greek State Scholarship Foundation, and A. G. by an MRC studentship.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 15, 1988; revised February 7, 1989.


Characterization of a murine homeo box gene, \textit{Hox-2.6}, related to the \textit{Drosophila Deformed} gene

Anthony Graham, Nancy Papalopulu, Joanne Lorimer, John H. McVey,Edward G.D. Tuddenham, and Robb Krumlauf

Laboratory of Molecular Embryology, National Institute for Medical Research, London NW7 1AA, UK
Characterization of a murine homeo box gene, *Hox-2.6*, related to the *Drosophila Deformed* gene

Anthony Graham, Nancy Papalopulu, Joanne Lorimer, John H. McVey, Edward G.D. Tuddenham, and Robb Krumlauf

Laboratory of Molecular Embryology, National Institute for Medical Research, London NW7 1AA, UK

The *Hox-2* locus on chromosome 11 represents one of the major clusters of homeo-box-containing genes in the mouse. We have identified two new members (*Hox-2.6* and *Hox-2.7*), which form part of this cluster of seven linked genes, and it appears that the *Hox-2* locus is related by duplication and divergence to at least one other mouse homeo box cluster, *Hox-1*. The *Hox-2.6* gene encodes a predicted protein of 250 amino acids, which displays extensive similarity in multiple regions to certain mouse, human, *Xenopus*, and zebra fish homeo domain proteins. The *Drosophila Deformed* (*Dfd*) gene also shares these same regions of similarity, and based on this sequence conservation, we suggest that *Hox-2.6* forms part of a vertebrate ‘*Dfd*-like’ family. *Hox-2.6* is expressed in fetal and adult tissues and is modulated during the differentiation of F9 teratocarcinoma stem cells. In situ hybridization analysis of mouse embryos shows that the *Hox-2.6* is expressed in ectodermal derivatives: spinal cord, hindbrain, dorsal root ganglia, and the Xth cranial ganglia. In the central nervous system, expression is observed in the most posterior parts of the spinal cord, with the anterior limit residing in a region of the hindbrain and no expression in the mid- or forebrain. In mesodermal structures, *Hox-2.6* is expressed in the kidney, the mesenchyme of the stomach and lung, and the longitudinal muscle layer of the gut. Expression has not been observed in derivatives of embryonic endoderm. The patterns of *Hox-2.6* expression in both mesoderm and ectoderm are spatially restricted and may reflect a role for the gene in the response to or establishment of positional cues in the embryo.

*Key Words: Hox-2.6; homeo box; mouse embryo; Drosophila Deformed; in situ hybridization; teratocarcinoma cells*

Received June 3, 1988; revised version accepted August 25, 1988.

The homeo box was first discovered in *Drosophila* in genes that were implicated in controlling pattern formation by specifying segmental pattern and segment identity (for review, see Gehring 1987). This process involves the determination of positional values along the anteroposterior body axis. More recently, it has been shown that several *Drosophila* homeo-box-containing genes are expressed in pattern that suggest roles in positional specification, independent of the process of segmentation (Doyle et al. 1986; Hoey et al. 1986; Akam 1987; Gehring 1987).

The homeo box can encode a 61-amino-acid protein domain that has a putative helix-tum-helix motif. This motif is implicated in DNA-binding activity (Desplan et al. 1985; Fainsod et al. 1986; Cho et al. 1988), and it is possible that the biochemical function of the homeo domain is to direct DNA binding and potentially regulate transcription. Homeo-box-containing genes have been identified and cloned from the genomes of widely sepa-

1 Present address: Haemostasis Research Group, Clinical Research Centre, Middlesex HA1 3UJ UK.
with two vertebrate genes not only in the homeo domain but also in other regions of the protein sequence. Information of this type suggests that vertebrate analogs of previously characterized Drosophila genes may exist. Therefore, it is important to determine and compare the coding information of more vertebrate homeo-box-containing genes to allow definition of closely related homeo box gene subfamilies, not based solely on the homeo box.

Whether conservation of structure reflects homologous functions for homeo box genes in the development of different species has yet to be shown. However, in view of the difficulty in defining mammalian genes responsible for regulating important developmental processes, they provide an opportunity to study molecules that are potentially involved in control during embryogenesis. The patterns of expression of mammalian homeo box genes during development are consistent with such a role (for review, see Feinberg et al. 1987). A number of groups have analyzed the expression of murine homeo box genes and found them to be temporally and spatially restricted in their patterns of expression (Gaunt et al. 1986; Dony and Gruss 1987). The murine homeo box genes appear to be expressed in similar but not identical overlapping domains, which often display differences in their anterior–posterior, dorsal–ventral, and medial–lateral levels of expression (Feinberg et al. 1987; Krumlauf et al. 1988; Gaunt 1988; Holland and Hogan 1988). The complex and overlapping patterns of transcription suggest that these genes may be involved in several different processes during embryogenesis, including the setup of positional cues.

The fact that vertebrate homeo box genes are found in similar types of clusters in all organisms so far analyzed is intriguing and may be a consequence of their modes of regulation or function. Thus, it is important to extend our understanding of the organization and expression of genes within these mouse clusters, to examine evolutionary or functional relationships between members of the homeo box gene family. This paper presents a further characterization of the Hox-2 locus, identifying two new genes (Hox-2.6 and Hox-2.7) and shows that there are at least seven linked members of this cluster. Analysis of the structure and expression of one of the new members, Hox-2.6, is detailed. The Hox-2.6 gene is expressed in a temporally and spatially restricted pattern during embryogenesis, and its predicted protein displays extensive identity to other homeo domain proteins. Based on multiple regions of similarity to the Drosophila Dfd gene, Hox-2.6 can be seen to be a member of a 'Dfd-like' family of vertebrate genes.

Results

Hox-2.6 gene structure and protein sequence

A number of overlapping genomic clones, which span the Hox-2 locus on chromosome 11, have been isolated from a mouse cosmid DNA library. Initially, a 220-bp fragment containing the Antp homeo box was used as a probe to isolate clones that span part of the locus. From these clones we were able to characterize the structure, sequence, and patterns of expression for one member of the locus, Hox-2.1 (Krumlauf et al. 1987). Subsequent genomic walking yielded more cosmid clones extending 5' and 3' of this region. Restriction mapping, sequencing, and hybridization analysis, using mouse and Drosophila homeo box probes, allowed construction of a map for the Hox-2 locus that positions seven homeo boxes over 80 kb (Fig. 1). This map is in agreement with previous results assigning five genes, Hox-2.1 to Hox-2.5, to the locus (Hart et al. 1985, 1987b) and, in addition, identifies two new members of the locus, which we term Hox-2.6 and Hox-2.7. The Hox-2.6 and Hox-2.7 genes are, at present, the 3'-most members of the locus and lie ~13 kb and 31 kb downstream of Hox-2.1.

A 600-bp Hox-2.6 probe containing part of the homeo box and 3'-flanking sequences was generated from a subcloned 4.4-kb BamHI fragment (BglII–BamHI; probe 1, Fig. 1). This probe was used to isolate a number of Hox-2.6 cDNA clones from a λgt10 library, prepared from 8.5-day p.c. mouse embryo RNA (Fahrner et al. 1987). The two longest cDNA clones (2 and 7, Figs. 1 and 2) and genomic clones from the Hox-2.6 region were sequenced. Sequence analysis reveals that the cDNAs contain a long open reading frame encoded by two exons, which in the genomic DNA are separated by a 790-bp intron, and contain consensus splice donor-acceptor sites. Hox-2.6 is transcribed in the same direction as Hox-2.1, and it appears that all seven genes in the locus have the same 5'→3' orientation (Hart et al. 1987b). Figure 2 shows complete sequence of the two cDNAs and some additional 3' sequence from a genomic subclone. The size of the major Hox-2.6 transcript, estimated by Northern RNA analysis (see Fig. 4), is 2.4 kb, and the cDNA clones account for only 1300 bp. The 3' ends of these clones have no adjacent polyadenylation signal and appear to have arisen by specific internal priming from an A-rich region between position 998 and 1104 (Fig. 2). There is a potential polyadenylation signal (AAATAA) at position 890 in the 3' untranslated region; however, RNase protection experiments have indicated that this site is not used (data not shown). Genomic probes 3' of the cDNA clones detect mRNA, suggesting that most of the sequence missing from our clones is derived from the 3'-untranslated end of the mRNA. Also contained within the 3'-untranslated region of Hox-2.6 is an AUUUA motif implicated in modulating the level of mRNA stability (Shaw and Kamen 1986).

Conceptual translation of the sequence in Figure 2 reveals that both cDNAs contain a long open reading frame. The first AUG, 36 bp after an in-frame stop codon, has been denoted the methionine-initiator codon. The predicted protein contains 250 amino acids, corresponding to a predicted molecular weight of 27,500, with an in-frame hexapeptide (Krumlauf et al. 1987) and homeo domain in the carboxy-terminal portion of the protein. Like Hox-2.1 and many other homeo-domain-containing proteins, Hox-2.6 is rich in prolines (18%), including a run of 14 consecutive prolines in the amino-terminal region.
Evolutionary relationships to the Hox-2.6 protein—a Dfd family

Sequence comparisons of the Hox-2.6 homeo domain with other homeo domain sequences are summarized in Table 1. The Hox-2.6 homeo domain shows 83% amino acid identity to that of Antp (McGinnis et al. 1984) and is therefore a member of the ANT-C/BX-C class of homeo domain. Considering murine homeo domains, the Hox-2.6 homeo domains show the closest similarity to those of Hox-1.4 and Hox-5.1, both with 93% amino acid identity (Duboule et al. 1986; Featherstone et al. 1988). In addition, sequences immediately adjacent to the Hox-1.4 and Hox-5.1 homeo domains, plus the hexa-peptides and amino-terminal regions, show extended region of identity with Hox-2.6 (Featherstone et al. 1988; B. Galliot, pers. comm.). These genes appear to form a subfamily of related genes.

Significant similarities to Hox-2.6 are also found in other vertebrate homeo-domain-containing proteins. These are the human HHO.c13 (Mavillo et al. 1986), the Xenopus Xho 1A (Harvey et al. 1986), and the zebra fish ZF 13 (Njolstad et al. 1988) genes. Both HHO.c13 and Xho 1A show 95% amino acid identity within the homeo domain, whereas ZF 13 shows 98% amino acid identity. All three proteins exhibit a high degree of identity in sequences immediately flanking the homeo domain, with those of ZF 13 being closest to Hox-2.6 (Njolstad et al. 1988).

An alignment comparison between Hox-2.6, Xho 1A, and HHO.c13 over the entire protein shows that sequences in addition to the homeo domain are highly conserved in these proteins [Fig. 3A]. The three proteins are different in size (232–255 amino acids), but if positioned for maximal alignment, one can see substantial conservation in the amino terminal region. This is more extensive in the case of Xho 1A than in the case of HHO.c13; yet in both instances, the first 30–35 amino acids show a high degree of sequence similarity. Within these amino acids, many of the differences represent conservative changes, and there is a block of 12 amino acids that are identical in all three proteins. The other region of notable similarity in these proteins begins around the hexapeptide and extends through the homeo domain and the first five to six amino acids on its carboxy-terminal side. The remaining regions of these proteins are highly divergent.

It was noted previously that HHO.c13 and Xho 1A shared related regions with the Drosophila homeo-domain-containing protein Dfd (Regulski et al. 1987). We therefore compared Hox-2.6 with Dfd, and Figure 3B diagrams blocks of sequence similarity identified with an optimal alignment between the Hox-2.6 and Dfd proteins. The Dfd protein is much larger than Hox-2.6 (586 vs. 250 amino acids). This difference appears to be the consequence of several regions rich or repetitive in single amino acids, such as Gly, His, Gln, or Asn, which are not present in the mouse gene. The Hox-2.6 homeo domain has 89% amino acid identity with the Dfd homeo domain. The overall amino acid identity to Dfd is 40% (102/250 amino acids) and is more extensive
Figure 2. Nucleic acid and protein sequence of Hox-2.6 from cDNA and genomic clones. The longest open reading frame and its conceptual translation are shown, and numbering starts on the first AUG 36 bp downstream of an in-frame stop codon. The intron is indicated by the large arrowhead. The beginning and end of the two cDNA clones [2 and 7], shown in Fig. 1, are marked by small arrowheads and their respective numbers. The open boxes outline the hexapeptide and the homeo domain.

Hox-2.6 shares almost exactly the same pattern of similarity with Dfd as it does with Xenopus 1A and HHO.c13. Several domains, including the highly conserved amino-terminal sequences in Hox-2.6, which are shared with the Xenopus and human genes, are also present in the Dfd protein. However the N-terminal identities in Dfd are partially disrupted by some of the repetitive amino acid regions [see Fig. 3B]. All of these genes also share many of the same amino acid sequence differences in the homeo domain when compared to the Antp homeo domain. Therefore, based on the multiple regions of homology it appears that Hox-2.6, HHO.c13, Xenopus 1A, Hox-5.1, Hox-1.4, and ZF 13 are all members of a vertebrate Dfd-like family related by a common progenitor that evolved before the divergence between the lineages that give rise to vertebrates and arthropods.

The Hox-2 and Hox-1 clusters are related by duplication

To extend our analysis of the Hox-2 locus further, we have isolated the Hox-2.7 gene and sequenced the homeo domain (Fig. 4). It shares only 68% amino acid identity (42/61) with the Antp homeo domain and, among other Drosophila homeo domains, is most closely related to S60Z1 [Doyle et al. 1986]. The predicted sequence in the homeo domain varies in only two positions with mhl9, also assigned to chromosome 11 [Lonai et al. 1987], and may represent the same gene. When compared with mouse homeo domains, Hox-2.7 displays the strongest identity with the Hox-1.5 domain and flanking regions (Fig. 4). This represents another example [like Hox-2.6] of a gene in the Hox-2 cluster, apparently more highly related to a gene in the Hox-1 cluster than to other Hox-2 members. A comparison of the sequence and organization of the seven genes in the Hox-2 cluster with the Hox-1 cluster sheds some light on the possible genealogy of these murine homeo box genes. The genes in both clusters are transcribed in the same 5'→3' orientation, and based on intergenic distances and sequence comparisons between members of the Hox-1 [Duboule et al. 1986] and Hox-2 clusters, it is possible to align some of the genes in these loci as shown in Figure 5.

Table 1. Homeo domain homologies to Hox-2.6

<table>
<thead>
<tr>
<th>Gene</th>
<th>Homology</th>
<th>Species</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antp</td>
<td>82%</td>
<td>Drosophila</td>
<td>McGinnis et al. (1984)</td>
</tr>
<tr>
<td>Dfd</td>
<td>89%</td>
<td>Drosophila</td>
<td>Regulski et al. (1987)</td>
</tr>
<tr>
<td>ZF 13</td>
<td>98%</td>
<td>zebra fish</td>
<td>P.R. Njostad et al. (pers. comm.)</td>
</tr>
<tr>
<td>Xho1A</td>
<td>95%</td>
<td>Xenopus</td>
<td>Harvey et al. (1986)</td>
</tr>
<tr>
<td>HHO.c13</td>
<td>93%</td>
<td>human</td>
<td>Mavillo et al. (1986)</td>
</tr>
<tr>
<td>Hox-1.4</td>
<td>93%</td>
<td>mouse</td>
<td>Duboule et al. (1986)</td>
</tr>
<tr>
<td>Hox-5.1</td>
<td>93%</td>
<td>mouse</td>
<td>Featherstone et al. (1988)</td>
</tr>
<tr>
<td>Hox-2.1</td>
<td>89%</td>
<td>mouse</td>
<td>Krumlauf et al. (1987)</td>
</tr>
<tr>
<td>Hox-2.3</td>
<td>82%</td>
<td>mouse</td>
<td>Meijlink et al. (1988)</td>
</tr>
<tr>
<td>Hox-2.7</td>
<td>75%</td>
<td>mouse</td>
<td>Fig. 4 (this paper)</td>
</tr>
</tbody>
</table>
Figure 3. Evolutionary relationships between Hox-2.6 and homeo domain proteins of other species. (A) Comparison of a related subclass of vertebrate homeo domain proteins, including mouse (Hox-2.6), human (HHO.cl4; Mavillo et al. 1986), and Xenopus (Xhox 1A; Harvey et al. 1986). The boxed region represents the homeo domain, and dashes indicate amino acid identity. (B) Diagrammatic representation of the relationship between Hox-2.6 and Drosophila Dfd. The shaded areas represent amino acid identity, and dashed lines demark related areas in the two proteins. (C) Generalized structure showing characteristic features and conserved regions based on comparison of several vertebrate homeo domain proteins. [AA] amino acid.

In cases where only the homeo domain and flanking sequences have been determined, comparisons reveal that the pairs Hox-2.7–Hox-1.5 and Hox-2.2–Hox-1.2 (Colberg-Poley et al. 1985b; Hart et al. 1987a) display the highest levels of sequence identity with each other, sharing in common specific amino acid changes when compared with the Antp homeo domain. The Hox-2.6 gene shows the strongest similarity with the Hox-1.4 homeo domain and surrounding sequences (as described in the previous section), and several other regions of the two proteins are also related [D. Duboule, pers. comm.]. Comparisons of other members of these clusters over the entire predicted protein sequences show that Hox-2.1–Hox-1.3 (Krumlauf et al. 1987; Odenwald et al. 1987; Fibi et al. 1988) and Hox-2.3–Hox-1.1 (Kessel et al. 1987; Meijink et al. 1987) are closely related to each other. The related region spans the hexapeptide region, the homeo domain plus surrounding sequence, and a
large region [30–50 amino acids] at the amino-terminal end of the proteins [Fibi et al. 1988; Papalopulu et al. 1988; J. Garbern et al., unpubl.]. Taken together, the data demonstrate that at least five genes in Hox-2 are related to five genes in Hox-1 via duplication and divergence of an ancestral cluster (Fig. 5). No mouse counterparts have been found as yet for the Hox-1.6 gene in the Hox-2 cluster, or for the Hox-2.4–Hox-2.5 genes in the Hox-1 cluster. However, it seems likely that further characterization of the overlapping regions of these clusters will reveal more related members and define the limits of the duplication.

**Developmental expression of Hox-2.6**

To begin to define the pattern of Hox-2.6 expression in the embryo and adult, we analyzed poly(A) + RNA extracted from teratocarcinoma cell lines and tissues at various stages by Northern hybridization. In view of the extensive sequence identity between Hox-2.6 and other subgroup members, such as Hox-1.4 and Hox-5.1, it was important to select probes that were specific for the Hox-2.6 gene. The carboxy-terminal region of these related proteins and the 3'-untranslated regions are highly diverged; therefore, the probe used was a single-stranded antisense RNA from this region of Hox-2.6 (Fig. 1, probe 1). Although it does contain the last 15 amino acids of the homeo domain, using high-stringency washing conditions with RNase A treatment (see Methods; Krumlauf et al. 1987), this probe does not detect transcripts or cloned DNA from the related genes [Hox-1.4 and Hox-5.1] in control experiments and is specific for Hox-2.6.

In tissues expressing Hox-2.6, the probe detects at least six transcripts ranging from 2.4 to 10 kb, including two major species at 2.4 and 4 kb [Fig. 6A]. The complex pattern of transcription reflects multiple transcripts from the Hox-2.6 gene and is analogous to results that we obtained with another member of the locus, Hox-2.1 (Krumlauf et al. 1987). The direct comparison of these adjacent genes in the Hox-2 cluster [Fig. 6] shows that they both have multiple transcripts that differ in size and intensity from each other. The bands do not appear to be unprocessed precursors, because intron probes do not hybridize to the multiple transcripts. It is possible that they represent RNAs capable of encoding multiple proteins from the same gene, as has been observed for a Xenopus homeo box gene (Cho et al. 1988).

The major tissues that express Hox-2.6 are the lung, spinal cord, and kidney, and a selected time course of expression in these tissues is shown in Figure 6A. The expression in lung is observed in all fetal, neonatal, and adult stages, and we have also detected transcripts by Northern analysis as early as 14.5 days p.c. The levels or complexity of expression in the lung do not change with time, but the 5-kb transcript is reduced in abundance when compared with other tissues. In the spinal cord, Hox-2.6 is expressed at the same levels in 14.5-day and 16.5-day embryos and is also present in the 18.5-day embryo, but we have not examined expression in the adult spinal cord. In contrast to the constant relative levels of Hox-2.6 mRNA in the lung and spinal cord during development, expression in the kidney undergoes a 10-fold drop between neonatal [1 week] and adult [6 weeks] stages. This same pattern of kidney expression is observed with the Hox-2.1 gene [Krumlauf et al. 1987]. Hox-2.1 is expressed in the adult kidney [Jackson et al. 1985] [although not observable on the exposure shown in Fig. 6B] at greatly reduced levels when compared with fetal and neonatal stages. Hox-2.6 is also expressed in adult testes and fetal liver, stomach, and gut but not in adult heart, liver, or spleen [data not shown]. In general, it seems that Hox-2.6 is expressed at a higher level in most tissues and developmental stages examined relative to Hox-2.1, with the exception of spinal cord. Both Hox-2.6 and Hox-2.1 show a reduction at the level of high molecular transcripts in the lung, suggesting tissue-specific differences in transcription or processing of these complex RNA species.

Based on the isolation of cDNA clones from an 8.5-day p.c. library, Hox-2.6 is expressed in 8.5-day p.c. embryos. It is difficult to obtain sufficient material to examine
earlier stages by Northern blots, but there have been a number of reports describing patterns of expression of murine homeo box genes in the F9 embryonal carcinoma cell line (Colberg-Poley et al. 1985a,b; Breier et al. 1986; Baron et al. 1987; Kessel et al. 1987; Rubin et al. 1987). The F9 cells are able to differentiate in vitro in response to chemical inducers to form derivatives of the extraembryonic endoderm, and we have used them to determine whether Hox-2.6 might be active in early stages of embryonic differentiation. Expression of Hox-2.6 was analyzed, by Northern blots, in stem cells and at 1, 3, and 5 days after retinoic-acid-induced differentiation [Fig. 6A]. Hox-2.6 transcripts are present in undifferentiated F9 stem cells at low levels. Transcript levels are increased after 1 day of differentiation to near maximal amounts and maintained at this high level throughout the remainder of the differentiation time course. This is in direct contrast to Hox-2.1, which is undetectable in F9 stem cells and accumulates to high levels after 1 day of induction [Fig. 6B]. This accumulation is transitory, and levels continually fall during further differentiation stages to very low levels by day 5. Treatment of F9 stem cells with the protein synthesis inhibitor cycloheximide also induced Hox-2.6 and Hox-2.1 RNA levels [Fig. 6]. In the case of Hox-2.1, this induction is not mediated by transcription but reflects mRNA instability involving the AUUUA motif [Shaw and Kamen 1986] in the 3'-untranslated region of the Hox-2.1 mRNA (R. Krumlauf et al., in prep.). Similar sequences reside in the 3' end of the Hox-2.6 mRNA [Fig. 2], and cycloheximide induction suggests that Hox-2.6 transcripts are also unstable in F9 stem cells. Thus, both Hox-2.1 and Hox-2.6 show an increase in expression following cycloheximide treatment and initial differentiation of F9 stem cells. However, the kinetics or temporal pattern of their response differs and depends on the de-
gree of differentiation of the F9 cells. In summary, the Northern analysis shows that Hox-2.6 and Hox-2.7 display similar overlapping patterns that only differ in the timing or level of expression in cell lines and a number of mouse tissues.

Hox-2.6—in situ hybridization

In situ hybridization has been used to extend and complement the analysis of expression by Northern blots that described the general patterns of Hox-2.6 expression. Antisense probes (probes 1 and 2, Fig. 1) and a sense probe were hybridized to sections of 12.5- and 14.5-day mouse embryos. We chose these stages, as they represent the stages of maximal expression of Hox-2.6. The sense probe acts as a control for nonspecific probe ‘stickiness’ to the sections. This probe did not highlight any features of the 12.5-day mouse embryo except red blood cells.

Of the three germ layers of the embryo, Hox-2.6 is expressed in derivatives of mesoderm and ectoderm but not in derivatives of endoderm. More specifically, Hox-2.6 is expressed in the following ectodermal derivatives: spinal cord, hindbrain, and some ganglia of the peripheral nervous system (Fig. 7). Hox-2.6 is not detected in the forebrain or midbrain, hence, expression within the central nervous system is restricted to more posterior regions. Transcripts are detected all along the spinal cord, from posterior portions (Fig. 7A) into the hindbrain (Fig. 7B,D). The anterior limit of expression maps to the hindbrain and does not seem to correlate with any obvious morphological change. This is clearly observed at 14.5 days, where one can see expression of Hox-2.6 extending along the spinal cord past the cervical flexure into the hindbrain (Fig. 7B). There appears to be no posterior limit of expression, and Hox-2.6 can be detected in the most posterior portions of the CNS. In contrast to Hox-2.1 and other homeo box genes that have been analyzed by in situ hybridization, Hox-2.6 is expressed uniformly in transverse sections of the spinal cord (Auwulewitsch et al. 1986; Toth et al. 1987; Holland and Hogan 1988) (Fig. 7C). It shows no dorsoventral or mediolateral spatial restriction at 12.5 days. Hox-2.6 is expressed in the peripheral nervous system in dorsal root ganglia and at very high levels in a ganglion that is probably the nodose ganglion of the Xth cranial nerve (Graham et al. 1988). Figure 7D shows expression of the gene in the hindbrain, in a dorsal root ganglion, and in the nodose ganglion. Expression of Hox-2.6 has not been observed in other cranial ganglia. The CNS is derived from the neural tube, and the peripheral nervous system has a neural crest origin. Thus, Hox-2.6 is expressed in both neural tube and neural crest derivatives.

Expression of Hox-2.6 in mesodermal structures is also spatially restricted. Hox-2.6 is expressed in the mesenchyme of the stomach and the lung, in the kidney, and in the longitudinal muscle layer of the gut. Figure 8 shows clearly the restriction of the expression of this gene to mesodermal tissues and not to endodermal tissue. This is obvious in the lung at 12.5 and 14.5 days, where the inner endodermal epithelial layer of cells is not expressing and the outer mesenchymal layer of cells is expressing Hox-2.6 (Fig. 8A,B). This mesodermal—endodermal restriction can also be seen in the stomach (Fig. 8C), whereas we can observe Hox-2.6 transcripts in the muscle layer of the gut in Figure 8D. The restriction of expression to the longitudinal muscle and not the circular muscle is more obvious at 14.5 days (Fig. 8D). Expression of Hox-2.6 has not been observed in other mesodermal organs such as the heart, the limbs, or those derived from somites. Those tissues that do express Hox-2.6 are derived from intermediate and lateral plate mesoderm.

This overall pattern of expression is very similar to the one observed for Hox-2.1 (Holland and Hogan 1988). One exception is the extent of expression that each gene shows along the rostral caudal axis. Most homeo box genes examined by in situ hybridization are expressed in the CNS; however the domains are overlapping but not identical and often display differences in the anterior limits of expression [Gaunt et al. 1988]. Figure 7 shows near adjacent sections of the hindbrain region of a 12.5-day mouse embryo that have been probed for either Hox-2.6 or Hox-2.1 (the Hox-2.1 probe is probe 1 of Krumlauf et al. 1987). One can clearly see that the expression of Hox-2.6 extends more anteriorly than that of Hox-2.1. The limits of expression of either gene do not seem to correspond to any clear morphological or biological cues at this embryonic stage. Thus, Hox-2.6 shows similar but subtly different patterns of expression to those of Hox-2.1 and, in common with all other homeo box genes, is spatially restricted in its patterns of expression during development. In view of the structural similarity between Hox-2.6 and other members of the Dfd-like subgroup, Hox-1.4 and Hox-5.1, it will be very interesting to determine whether their patterns of expression are also highly related.

Discussion

In this paper we present data relating to the characterization of a new murine homeo box gene, Hox-2.6. This gene and Hox-2.7, another new gene we have identified, form part of a group of seven tightly linked homeo box genes in the Hox-2 cluster located on mouse chromosome 11 (Fig. 1). To date, nearly 25 murine homeo box-containing genes have been identified. Therefore, the Hox-2 locus represents part of a large gene family related by at least one common element, the homeo box. Sequence analysis reveals a structure for Hox-2.6 that is similar to that of many other murine and vertebrate homeo box genes. Typically, Hox-2.6 is comprised of two exons separated by an intron of about 1 kb (790 bp). The protein-coding information (250 amino acids) is located in the 5' portion of the mRNA, which has a long 3'-untranslated region. A conserved hexapeptide is located close to the homeo box but on a separate exon, with the splice site located 15–40 bp upstream of the homeo box. The predicted proteins are very rich in prolines and serines, and the distribution of these prolines
throughout the non-homeo-domain portion of the protein should generate a nonhelical structure. Some of these features are summarized in Figure 3C, which presents a generalized structure for these proteins.

Hox-2.6 was found to be related to other mouse homeo-domain proteins (Hox-1.4 and Hox-5.1), based on sequence identity in several domains of the protein. This pattern of identity between different mouse genes appears to be the result of duplication and divergence of an ancestral homeo box cluster. Sequence comparisons of the Hox-1 and Hox-2 clusters reveal that at least five members of each cluster are highly related to members in the opposite cluster, suggesting a duplication event (Fig. 4). The relationship of other homeo box clusters to Hox-1 and Hox-2 is not well established. However, based on the identity between Hox-2.6 and Hox-5.1, some other genes in the Hox-3 and Hox-5 clusters may also be related by duplication to Hox-1 and Hox-2. Consequently, some mouse genes may form a subgroup with as many as four highly related members.

Figure 7. Expression of Hox-2.6 in ectodermal derivatives, as shown by in situ hybridization to embryo sections. (A) Hox-2.6 expression in posterior spinal cord of a 12.5-day p.c. embryo. (B) Expression of Hox-2.6 at the junction of the hindbrain and spinal cord of a 14.5-day p.c. embryo. (C) Transverse section of thoracic spinal cord of a 12.5-day p.c. embryo probed for Hox-2.6. (D) Expression of Hox-2.6 in hindbrain and constituents of the peripheral nervous system of a 12.5-day p.c. embryo. Bright-field and dark-ground pictures are shown. (RBC) Red blood cells; (SC) spinal cord; (HB) hindbrain; (DRG) dorsal root ganglion; (NG) nodose ganglion; (PV) pre-vertebrae.
Hox-2.6 structure and expression

Figure 8. Expression of Hox-2.6 in mesodermal derivatives, as shown by in situ hybridization, to embryo sections. (A) Hox-2.6 transcripts in the lung mesenchyme in transverse sections of 12.5-day p.c. embryo. (B) Lung of a 14.5-day p.c. embryo probed for Hox-2.6. (C) Hox-2.6 expression in the stomach mesenchyme of a 12.5-day p.c. embryo. (D) Expression of Hox-2.6 in the longitudinal muscle of a 14.5-day p.c. embryo. Bright-field and dark-ground pictures are shown. [E] Endoderm; [M] mesenchyme; [PV] pre-vertebrae; [LM] longitudinal muscle; [CM] circular muscle.

In a larger context, one can note relationships between a number of homeo box genes from a range of different organisms. Hox-2.6 not only shows similarity to two mouse genes [Hox-1.4 and Hox-5.1] but also to a human gene, HHO.c13 (Mavillo et al. 1986), a Xenopus gene, Xhox 1A (Harvey et al. 1986), a zebra fish gene, ZF 13 (Njolstad et al. 1988), and the Drosophila gene, Dfd (Regulski et al. 1987). High levels of sequence identity are present in regions other than the homeo domain, including the hexapeptide and amino-terminal sequences.

We suggest that these genes represent a Dfd-like family, based on their multiple regions of similarity. However, in view of the fact that there appear to be at least three murine genes that all show similar degrees of identity to Dfd, it is clear that there is not a precise murine equivalent to the Dfd gene.

It may generally be possible to classify several of the nonlinked homeo box genes from different species into subgroups, based on conserved sequences in their amino-terminal region (referred to as the conserved
amino-terminal subfamily domain in Fig. 3C). This conservation in several species supports the idea that these domains may be functionally significant. Hox-2.1 forms another such subgroup with the zebra fish ZF 21 gene [P.R. Njolstad et al., pers. comm.] and the Xenopus XIH box 4 gene (Fritz and De Robertis 1988). Some of these subfamilies may also be related to other Drosophila homeo domain proteins in a manner analogous to Hox-2.6 and Dfd.

The Hox-2.6 gene, as analyzed by in situ hybridization and Northern blots, is expressed in fetal and adult tissues. Major sites of expression are the lung, kidney, central and peripheral nervous systems, stomach, gut, and testis. Within visceral organs, Hox-2.6 is expressed in derivatives of mesoderm but not endoderm. This is shown in Figure 8, which illustrates that expression is restricted to the mesoderm of the stomach, gut, and lung, and is not seen in the endodermal epithelium. Within ectodermal derivatives, Hox-2.6 is expressed in the CNS, in the spinal cord and hindbrain, and in the PNS, in the dorsal root ganglia, and in the nodose ganglia, as shown in Figure 7.

The pattern of Hox-2.6 expression that we observe in the 12.5-day embryo is complex and does not correspond to any obvious biological or morphological signal. It seems probable that Hox-2.6 is responding to anteroposterior positional cues but is not doing so in any simple or exclusive fashion. It is clear that two structures derived from the same germ layer and sharing the same anteroposterior position do not necessarily both express this gene. This is illustrated well in the case of the lung mesenchyme, which does express this gene, and the limbs, which do not. Both structures share a similar anteroposterior position but exhibit differential expression of this gene. Thus, if position is of importance in determining the patterns of Hox-2.6 expression, it is likely to have been modified by other cues to produce the observed pattern at 12.5 days. This is comparable to the patterns of expression that are observed in later stage Drosophila embryos, where the more simplistic earlier patterns of expression have been modified during development [Martinez-Arias et al. 1987]. Hence, a simple early pattern of expression in mouse embryos may be masked later in embryogenesis.

There are similar regions of expression shared between vertebrate and Drosophila homeo box genes in comparable stages of development. It has been shown that a number of Drosophila homeo box transcripts accumulate to their highest levels in the embryonic CNS (for review, see Doe and Scott 1988). Expression in the CNS is observed with a large number of homeo box genes from a wide range of species. Expression of homeo box genes in the CNS has been described for humans [Simeone et al. 1986], for Xenopus [Carrasco and Malacinski 1987; Condie and Harland 1987], and the zebra fish [Njolstad et al. 1988]. With regard to the mouse, we have shown that Hox-2.6 is expressed in the CNS, and this has also been demonstrated with Hox-1.2 [Toth et al. 1987], Hox-1.3 [Dony and Gruss 1987], Hox-1.4 [Toth et al. 1987; Wolgemuth et al. 1987], Hox-1.5 [Fainsod et al. 1987; Gaunt 1987], Hox-1.6 [Baron et al. 1987], Hox-2.5 [Feinberg et al. 1987], Hox-2.3 [Deschamps et al. 1987], Hox-2.1 [Holland and Hogan 1988], and En-1 and En-2 [Joyner and Martin 1987]. These genes are expressed in an overlapping pattern along the anteroposterior axis of the CNS, as illustrated for Hox-2.1 and Hox-2.6 in Figure 9, but they also exhibit subtle differences in their mediolateral and dorsoventral planes of expression. Hox-2.6 is expressed uniformly in cross sections of the spinal cord [Fig. 5C], yet Hox-2.1 is expressed in a dorsally restricted manner [Holland and Hogan 1988].

Homeo box genes are clearly implicated in determining the identity of cells in the Drosophila blastoderm, and recent studies would also suggest that they may act in an analogous manner in the Drosophila nervous system [Doe et al. 1988]. For example, it has been reported that ftz in Drosophila and mec-3 in Caenorhabditis elegans [Way and Chalfie 1988] are necessary for specification of neuronal identity. Thus, homeo box genes in vertebrates, as well as in invertebrates, may possibly act in controlling neuronal fate. The specification of neuronal fate is certainly a very complex system, which may require a family of proteins that can act in concert to define cell fate/position. The large size of the family of murine homeo box genes and the fact that they all seem to be expressed in the CNS make them good candidates for performing such a role.

The three Dfd-related murine genes, Hox-2.6, Hox-1.4, and Hox-5.1, may act potentially in an antagonistic or synergistic manner. It will be of great interest to compare their patterns of expression and regulation in more detail. The one feature that may be shared is the extent of expression of these genes in the ectoderm along the rostrocaudal axis at comparable stages of development. Dfd is expressed near the rostral end of the Drosophila embryo [Martinez-Arias et al. 1987], and we have shown that Hox-2.6 is expressed along the spinal cord into the hindbrain of the murine embryo. Hox-1.4 is also expressed in the murine hindbrain [Toth et al. 1987], as is Hox-5.1 [Featherstone et al. 1988]; and Northern analysis shows HH0.c13 transcripts in the brain but not in the forebrain [Mavillo et al. 1986]. We have also presented evidence illustrating that Hox-2.6 is expressed more anteriorly than the adjacent Hox-2.1 gene [Fig. 9], and it has been reported that Hox-2.5 is not expressed in the hindbrain but is restricted to the spinal cord [Feinberg et al. 1987]. Therefore, it would appear that there is a correlation between the anterior limit of expression of a gene in the ectoderm and the relative position within a cluster, as has also been observed in Drosophila [Akam 1987]. This correlation may also extend to the subfamily to which that given gene is a member.

Methods
Isolation of clones and sequencing
Cosmid clones were isolated from a mouse (129) genomic library, prepared with the vector pcos2EMBL (a gift from Anna-Marie Frischauf), by screening with subcloned probes from the Hox-2.1 gene [Krumlauf et al. 1987]. The clones were mapped
and flanking regions subcloned to be used for further screening of overlapping cosmids. Homeo box regions in the cosmids were identified by hybridization with a Drosophila Antp (provided by Walter Gehring) and mouse (Hox-2.1) homeo box probe at reduced stringency according to Holland and Hogan [1986]. Homeo box fragments were then subcloned and used to screen a mouse agt10 cDNA library [Fahrner et al. 1987] prepared from 8.5-day embryos for Hox-2.6 clones. All cosmid and cDNA subcloning was performed using the vector Bluescript pKSM13+ [Stratagene]. Double-stranded DNA from the cDNA and genomic clones was sequenced by the dideoxy chain-termination method, using the array of primers in the Bluescript vector and the Sequenase polymerase [US Biochemical Corporation] according to manufacturer's directions. All sequencing was performed in both directions using both the standard and dITP reactions to avoid the compressions in GC-rich regions.

**Cell culture**

F9 stem cells were maintained in Dulbecco’s modified eagle’s medium [DMEM], supplemented with 10% fetal calf serum on gelatinized plastic tissue culture dishes. The cells were induced to differentiate by adding retinoic acid [5 × 10^{-8} M], cAMP [10^{-4} M], and IMBX [10^{-4} M] to cultures, and the cells were maintained in this differentiation media for 1–5 days, with the media being changed every 2 days. In the case where stem cells were treated with the protein synthesis inhibitor cycloheximide, the inhibitor was added to the cultures to a final concentration of 75 μg/ml for a total of 6 hr before harvesting. Under these conditions in control experiments, 95% of protein synthesis is blocked and 60% of the cells can recover to reinitiate protein synthesis when the inhibitor is removed. Cells were harvested for RNA at various time points by washing twice with PBS and lysing them on the plate in LiCl urea.

**RNA isolation and Northern hybridization**

Poly[A]+ mRNA was isolated from F9 cells or mouse tissues essentially as described in Krumlauf et al. [1987]. Briefly, the samples were homogenized in 3 M LiCl–6 M urea on ice for 2
min and stored overnight at 4°C to precipitate the RNA. The RNA was harvested by centrifugation, redissolved in 10 mM Tris-HCl [pH 7.6], 1 mM EDTA, and 0.5% SDS and extracted several times with phenol : chloroform. Following ethanol precipitation, the poly[A]+ mRNA was isolated by oligo(dT) cellulose chromatography. The RNA samples were electrophoresed in denaturing formaldehyde–agarose gels, transferred to GeneScreen (Du Pont) in 20 x SSC, and coupled to the membrane by UV cross-linking and baking, as described in Krumlauf et al. [1987]. The filters were hybridized using 32P-labeled single-stranded antisense RNA probes produced by in vitro transcription of subcloned Hox-2.1 and Hox-2.6 cDNA clones with T7 or T3 RNA polymerase. Probes were selected from regions of genes that had a minimum identity with other subgroups to minimize any cross-hybridization. Hybridization conditions were 60% formamide, 5 x SSC, 0.1% bovine serum albumin, 0.1% Ficoll 400, 0.1% polyvinylpyrrolidone, 20 mM sodium phosphate [pH 6.8], 1% SDS, 7% dextran sulfate, 100 μg/ml tRNA, and 10 μg/ml poly(A) at 65°C for 16 hr. The filters were washed in 0.1 x SSC, 0.5% SDS at 80°C for 2 hr, and in the cases where high-stringency conditions were required to ensure that the signal was not derived from related genes, the filter was treated further with RNase A. The filters were incubated in 2 x SSC, 0.1% SDS at 52°C for 1 hr. Probe was removed from filters (not treated with RNase A) for rehybridization by washing in 70% formamide at 80°C for 30 min and reexposing the filter to ensure probe removal.

In situ hybridization

The protocol used was basically that of Wilkinson et al. [1987], with some modifications, and is essentially as follows. Mouse embryos were fixed in 4% paraformaldehyde in phosphate-buffered saline [PBS] overnight at 4°C, and then embedded in paraaffin wax. Six-micrometer sections were cut and dried onto gelatin-subbed slides. The sections were dewaxed in xylene, treated with 0.2 M HCl, refluxed in paraformaldehyde, subjected to proteinase K, and then treated with acetic anhydride. After dehydration, probes were redissolved at a final activity of 1.2 x 10^5 dpm/kb/p-l in hybridization buffer (50% formamide, 0.3 M NaCl, 20 mM Tris, 5 mM EDTA, 10 mM NaPO4, 10% dextran, 1 x Denhardt's tRNA). Hybridization was overnight at 50°C. The sections are washed at 65°C in 50% formamide, 2 x SSC, 100 mM dithiothreitol [DTT] for 40 min, followed by incubation with RNase A at 20 μg/ml in 0.5 M NaCl, 10 mM Tris, 5 mM EDTA for 30 min. Sections were washed again in 50% formamide, 2 x SSC, and 100 mM DTT for 30 min, followed by two 15-min washes in 2 x SSC and 0.1 x SSC. Sections were dehydrated through alcohol solutions containing 0.3 M NH4OAc. Slides were dipped in a mix of Ilford K5 nuclear emulsion and glycerol/water [6 ml emulsion in 8.2 ml H2O/0.18 ml glycerol] and kept at 4°C until developed. Exposures were between 10 and 12 days. Sections were stained in 0.02% toluidine blue for 1 min and then mounted in Permount.

Acknowledgments

We thank Dr. Bernhard Hermann for help in screening the cosmId libraries and isolating pCos 5.4 and 3.1, and Drs. Dennis Duboule, Mark Featherstone, Brigitte Galliot, Pal Njolstad, Jim Garbern, Ward Odenwald, and Edoardo Boncinelli for sharing unpublished sequence information and ideas about the organization and evolution of homeo box genes. We also thank Drs. David Wilkinson, Peter Holland, Frank Ruddle, Mike Snow, Andy McMahon, Dennis Summerbell, and Mai Har Sham for valuable discussions and comments on the manuscript. N.P. was supported by the Greek State Scholarship Foundation and A.G. by a Medical Research Council studentship.

References


