Genetic and Molecular Studies of the Dominant hemimelia Locus in the Mouse.

Maureen Higgins

Ph.D

University of Edinburgh

1991

EX BIBL UNIV. EDINBURGH

Declaration

I declare that,

(a) This thesis has been composed by myself, and

(b) The work is either my own or the author involved is clearly stated.

Acknowledgements

There are a great many people to whom I owe much thanks for all their help and encouragement over the last three years, each of whom contributed to making the last three years, if not always enjoyable, at least bearable!

John & Bob, for taking me on, letting me stay (!), and being great supervisors - thanks,

Chris, for all her help and friendship, particularly during the first six months,

Jean, Margaret & Roberta, for running a relaxed lab,

John I, for always having an answer to questions beginning with "What would happen if.....?",

Wendy & John M, for help with PFGE,

Paula Murphy, for always being herself,

Sinead, for being a great bench mate in more ways than one,

Paula Mon., for lateral thinking,

Maureen & Dennis, for running a well organised department,

SERC & Lloyds bank, for financial support,

Tom & Ted, for graphics & photographs,

Sheila & Helen, for running an excellent library service,

My family, especially Mum & Alan for the much needed trip to Wales,

Mairi & Alex, Fiona & Jim for always being great friends,

and finally Dave, who gave the most and asked the least back in return.

The work described in this thesis was carried out at the MRC Human Genetics Unit and the University Dept. of Obstetrics & Gynaecology. I am particularly grateful to Prof. Evans for financial support. This thesis is dedicated to the memory of my father.

Glossary of Terms

Hemimelia:	Condition of having defective limbs.
Preaxial:	Situated or occuring before an axis. In anatomical usage preaxial refers to the medial (tibial) aspect of the lower limb.
Luxation:	Dislocation.
Polydactyly:	Presence of more than the normal number of fingers or toes on a hand or foot.
Oligodactyly:	Deficiency of fingers or toes.
Triphalangy:	Deformity of hand/foot in which there are three phalanges in the thumb/big toe instead of the normal two.
Hallux:	Big toe.
Penetrance:	The frequency with which a gene manifests itself in the phenotype of carriers.
Expressivity:	Phenotypic expression of a penetrant gene and may be slight, intermediate or severe.

Abstract

Dominant hemimelia (*Dh*) is a semi-dominant mutation which arose spontaneously in 1954. *Dh* /+ mice are asplenic, have a slightly smaller stomach and digestive tract than their +/+ siblings. They also exhibit preaxial abnormalities of the hind limb which include triphalangy of the hallux, polydactyly, oligodactyly and a reduction or absence of the tibia. The number of ribs, sternabrae and pre-sacral vertebrae is reduced. The abnormalities are more severe in homozygous animals and are usually fatal.

The earliest morphological defect which can be detected is in the splanchnic mesoderm, a derivative of the lateral mesoderm. It has been hypothesised that *Dh* interferes with the normal structural arrangement of the cells of the splanchnic mesoderm and that all defects in *Dh* animals are traceable to this structure. In the introduction to this thesis it is proposed that the primary effect of the *Dh* mutation occurs during segmentation, i.e. somitogenesis. It is proposed that *Dh* disrupts the process of somitogenesis and results in the loss of the normal organised structure of the splanchnic mesoderm and in vertebral malformations. The pleiotropic phenotype of *Dh* is a direct consequence of the combined effect of these two abnormalities.

The observation that *En-1*, a homeobox containing gene, mapped close to or at the *Dh* locus prompted the work described in this thesis. The suggestion that homeobox containing genes are key genes in the control of development has been stengthened by several lines of evidence and provoked the question of whether the *Dh* mutation represented a mutant allele of *En-1*. The principle aim of this project was to determine whether *Dh* and *En-1* were alleles of one another and, if not, devise means of identifying and cloning the *Dh* gene.

Dh was found to map 0.7±0.4 cM proximal to *En-1* and 1.1±0.4 cM distal to *Emv-17* and is therefore unlikely to represent a mutant allele of *En-1*. The physical localisation of the *Dh* locus was attempted using pulsed field gel electrophoresis (PFGE). The PFGE map around *En-1*, which encompasses 1500kb, detected the presence of 5 CpG islands and therefore, by inference, genes. Two of these CpG islands are within 800kb 3' to *En-1*, the remaining three islands are situated within 600kb 5' to *En-1*. Whether these any of these islands are associated with the *Dh* locus will depend on the ratio of genetic to physical distance in this region of the genome. The localisation of the *Dh* locus and the orientation of the *En-1* gene relative to the chromosome was attempted via the mapping of the recombination breakpoint of the four animals recovered from the linkage analysis which were recombinant between *Dh* and *En-1*.

In order to clone the *Dh* locus a gene targeting vector was designed and constructed. This vector contains a mammalian and a yeast selectable marker. The presence of the yeast selectable marker will allow the construction of small easily manageable YAC (yeast artificial chromosomes) libraries, which can clone up to 2Mb, enriched for the *En-1* region of the genome. The combination of the PFGE mapping studies and the ability to clone large fragments of DNA provide a powerful means of specifically identifying and cloning the *Dh* locus.

The vector will also be used for gene targeting in pluripotent mouse embryo derived stem (ES) cells. This technique provides a way of determining the genetic function of *En-1* as it will replace the endogenous *En-1* gene with one which has been disrupted, and can therefore generate homozygous *En-1* null mutant mice.

Table of Contents

Chapter 1 Introduction **1**

1.1 Early Mouse Development 3
1.1.1 Gastrulation and mesoderm segmentation 10

1.2 Embryonic Induction 14
1.2.1 Timing and localisation of the response 19

1.3 The *Dh* Mutation 21

1.4 Homeobox-Containing Genes 25
1.4.1 The homeobox encodes a transcription factor 28
1.4.2 Evolutionary conservation of homeobox-containing genes 31
1.4.3 *Engrailed*: Role in *Drosophila* and mouse 33
1.4.4 Homeobox genes and induction 39

1.5 Aims of Thesis 42

Chapter 2 Materials and Methods **43**

2.1 Anaesthesia 44

2.2 Surgical Procedures 45
2.2.1 Tail Bleeding 45
2.2.2 Tail tip removal / isolation of high molecular weight DNA 46

2.3 Electrophoresis of *Peptidase-3* 46
2.3.1 Preparation of haemolysate 47
2.3.2 Electrophoresis 47

2.3.3 Detection of *Peptidase-3* 47

2.4 Statistical Analysis 48
2.4.1 Recombination frequency 48
2.4.2 Chi square tests 49

2.5 Bacterial Culture 49
2.5.1 Media used 49
2.5.2 Bacterial strains 50

2.6 Plasmid Cloning Vectors 51

2.7 Cosmid Library Screening 52

2.8 Manipulation of Plasmid DNA 53
2.8.1 Preparation of *E. coli* competent for DNA transformation 53
2.8.2 Bacterial transformation 53
2.8.3 Screening bacterial colonies 54
2.8.4 Plasmid / cosmid DNA minipreparation 54

2.8.5 Large scale plasmid DNA preparations 55

2.9 Preparation of Mammalian DNA 55
2.9.1 Preparation of high molecular weight DNA 55
2.9.2 Preparation of DNA agarose blocks 56
2.9.3 Preparation of yeast chromosome markers 57
2.9.4 Preparation of lambda bacteriophage concatemers 57

2.10 Manipulation of DNA by Enzymes 58
2.10.1 Restriction endonuclease digestion 58
2.10.2 Restriction endonuclease digestion of DNA agarose blocks 58
2.10.3 Ligation of DNA molecules 59
2.19.4 Filling In Cohesive Termini 59
2.10.5 Enzymatic phosphorylation of oligonucleotide linkers 59
2.10.6 Ligation of phosphorylated linkers to DNA 60

2.11 Gel Electrophoresis 60
2.11.1 Conventional gel elecltrophoresis 61
2.11.2 Pulsed field gel electrophoresis 61
2.11.3 Preparative gel electrophoresis 62

2.12 Southern Blotting 62

2.13 Preparation of Radiolabelled DNA Probes 63
2.13.1 Nick translation 63
2.13.2 Random prime labelling 64
2.13.3 End labelling of oligonucleotides 64

2.14 Hybridisation 65

Chapter 3 The *Dh* Phenotype 66

3.1 Introduction 67

3.2 Hind Limb Abnormalities 68

3.3 Asplenia 73

3.4 Kidney Abnormalities 74

3.5 Recovery of the *Dh* Mutation 76

3.6 Discussion 79

Chapter 4 Linkage Analysis 82

4.1 Introduction 83

4.2 Results 84
4.2.1 Generation and maintenanace of mouse stocks 87
4.2.2 Close linkage of *En-1* to the *Dh* locus on chromosome 1 89
4.2.3 Factors affecting recombination frequency 96

4.3 Discussion 98

Chapter 5 Long & Short Range Mapping Around The *En-1* Locus 105

5.1 Introduction 106

5.2 Results 109
5.2.1 Screening a cDNA library 109
5.2.2 Screening a cosmid library 109
5.2.3 Cosmid mapping 111
5.2.4 Isolation of a 5' mapping probe 116
5.2.5 Isolation of a 3' mapping probe 116
5.2.6 Pulsed field gel electrophoresis 118
5.2.7 Long range partial mapping 122
5.2.8 Long range mapping of *Dh* DNA 125

5.3 Discussion 126

Chapter 6 Construction of En-1 Targeting Vector 131

6.1 Introduction 132

6.2 Results 139
6.2.1 Modification of pBLRC plasmid 139
6.2.2 Generation of pBHI 22 plasmid 141
6.2.3 Production of pEN-TK1/3 141
6.2.4 Attempts to subclone the Neo-Leu cassette 144

6.3 Discussion 150

Chapter 7 Discussion 154

7.1 Summary of Results 155

7.2 Future Work 157

References 162

Abbreviations 190

Chapter One

Introduction

Introduction

The work presented in this thesis came about as a result of the finding that, in the mouse, the homeobox containing gene *Engrailed-1* (*En-1*) mapped in the vicinity of the developmental mutation *Dominant hemimelia* (*Dh*) (Hill *et al.*, 1987; Joyner & Martin, 1987). This was particularly provocative due to the fact that homeobox containing genes were known to play an important role in *Drosophila* development and raised the question of whether *Dh* was a mutant allele of *En-1*. *Dh* is a semi-dominant mutation the main features of which are asplenia, pre-axial abnormalities of the hind limb, decreased number of pre-sacral vertebrae and widespread abnormalities of the uro-genital and digestive systems. The abnormalities are more severe in the homozygote and are usually lethal. The mutation can be traced to 9.5 days where particular regions of the splanchnic mesoderm are either missing or defective and it has been hypothesised that the various aspects of the phenotype result from a failure of the normal inductive action of the splanchnic mesoderm (Green, 1967). Although the mutation can be traced to a defective splanchnic mesoderm no explanation was offered as to why this tissue was affected. In the introduction to this thesis it is postulated that the primary defect in *Dh* mice occurs during somitogenesis, i.e. segmentation, and that the defective splanchnic mesoderm occurs as a direct consequence of this. The pleiotropic phenotype seen in *Dh* mice then results from to the impairment / failure of later inductive interactions. Thus, before discussing the *Dh* phenotype, it is relevant to introduce early mouse development, with particular regard to the process of mesoderm segmentation, and the nature and importance of embryonic inductive interactions. As mentioned above, the initial finding that *En-1* mapped close to or ar the *Dh* locus suggested that these genes may be allelic. Therefore following a discussion of the *Dh* phenotype there is a review of homeobx containing genes with particular regard to *En-1*.

1.1 Early Mouse Development

The development of a complete organism from a single fertilised egg, a process which is usually flawless, is one of the most fascinating aspects of biology and one that occurs so efficiently and reliably that the complexity of the process is very often taken for granted (reviewed by Theiler, 1972; Snell & Stevens, 1975; Hogan *et al.*, 1986; Wilkins, 1986).

In the mouse embryonic and fetal growth and development takes 20 days from fertilisation to birth. The first phase of development up to the end of gastrulation takes 7days from the time of fertilisation. The period of organogenesis occurs from 7-14 days and the growth phase continues thereafter into adult life.

During the first 4.5 days the embryo exists in an independent unattached state undergoing cleavage. The first two divisions occur in the oviduct up to 24 hours post coitus (p.c.). The following divisions occur more rapidly, each division taking 5 to 10 minutes with 12 hour (hr) intervals. By 60 hrs p.c. (2.5days) the pre-implantation embryo, now known as the morula, has reached the 16 cell stage and typically journeys from the oviduct into the uterus some 6 to 12 hours later (Lewis & Wright, 1935). During the next day and a half the morula develops into a hollow, spherical, fluid filled ball of cells known as the blastocyst which then settles into the uterine crypts between 4.3 to 4.5 days (Eaton & Green, 1963). As the embryo is undergoing cleavage in the first few days of development two important events occur: compaction and the formation of two cell lineages.

Single blastomeres from 2 and 4 cell stage morulae have been shown to be able to give rise to normal mouse whereas single blastomeres from the early 8 cell stage do not have the ability to generate a whole mouse but can contribute to a wide range of tissues in chimaeric offspring (Kelly, 1977). As cleavage proceeds to the 16 cell stage

the developmental option of the cells is narrowed resulting in the generation of 2 distinct cell populations: an outer layer of trophectoderm and an inner layer of cells called the inner cell mass (ICM). The ICM of the late blastocyst contributes directly to the foetus and to certain extraembryonic membranes. The trophectoderm however contributes to tissues involved solely in the implantation process.

The differentiation into the two cell types begins at compaction which occurs at the 8 cell stage. The major effect of compaction is the transformation of the cells from spherical to columnar and the development of a variety of intercellular junctions (Dulcibella, 1977). Compaction is also regarded as playing a major role in determining polarity within the blastomeres of the 8 cell stage embryo. The blastomeres develop distinct apical and basal membrane and cytoplasmic domains. This cellular polarisation forms the basis of the polarisation hypothesis which suggests that the two distinct cell lineages of the blastocyst, i.e the trophectoderm and the ICM, differentiate as a result of differential inheritance of membrane and cytoplasmic molecules such as plasma membrane glycoproteins, receptors and cytoskeletal organelles (Pratt *et al.*, 1981; Johnson & Ziomek, 1981). However the segregation of information at the 16 cell stage is not thought to be irreversible as cells remain developmentally labile until 32 cell stage (Rossant & Frels, 1981). Polarization at compaction is therefore thought to guide cells towards their final fate rather than fixing them.

The final phase of pre-implantation development involves the ICM and trophectoderm differentiating to form a further two cell types each (see Fig1.1). The trophectoderm cells differentiate into the mural trophectoderm, which is in contact with the blastocoel cavity, and a cap of polar trophectoderm cells adjacent to the ICM. The mural trophectoderm cells eventually go to form the polytene giant cells of the implanting blastocyst. The ICM cells in contact with the blastocoel form the primitive endoderm which then surrounds a core of primitive ectoderm (epiblast) from which the

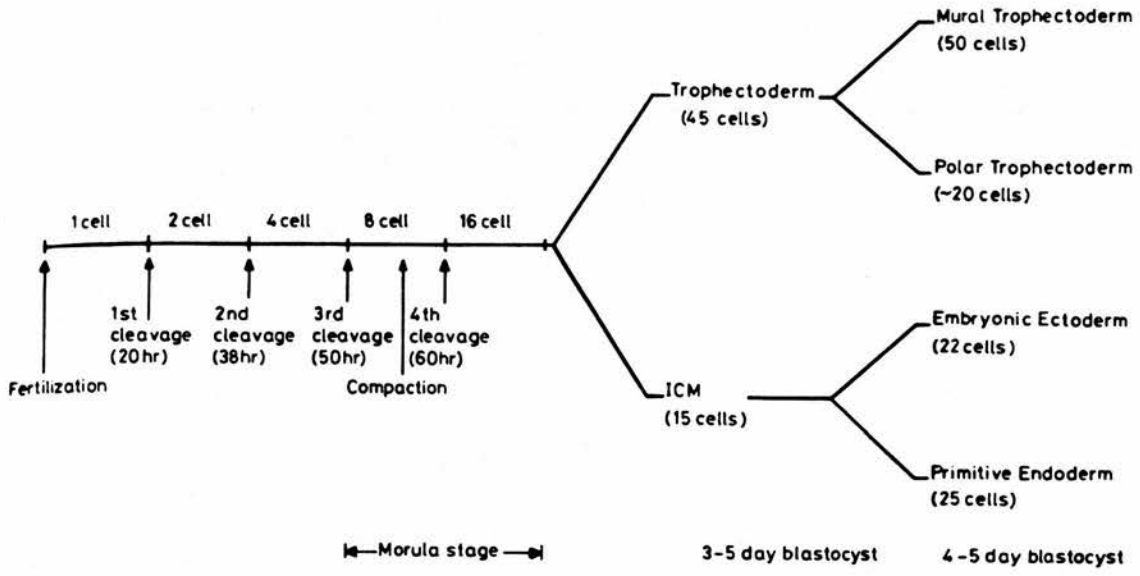


Figure 1.1. Time course and stages of mouse pre-implantation development (reproduced from Wilkins, 1986).

embryo proper develops. The trophoblast cells and the primitive ectoderm form the yolk sac, amnion and placenta.

At this late blastocyst stage (approximately 120 cell) 4.5 days p.c, implantation occurs and the major phase of embryonic development begins. During this time the general body plan is laid down, the 3 germ layers form, somitogenesis occurs followed by organogenesis and general foetal growth.

Up to the time of implantation there has been no growth in size in the embryo however the period post-implantation is one of rapid growth, e.g. by 7.5 days , 3 days after the 120 cell stage, the embryo proper consists of 10^5 cells (McLaren, 1976), the rate of cell division and development increases further with the establishment of the placental connections. The mesoderm originates at 6.5 days with the appearance of the primitive streak, which is a thickened region at the posterior end of the embryonic ectoderm, and from which all major organ systems develop.

However one of the first events post-implantation is the enlargement of the blastocoele followed by the growth of the ICM downward into the enlarged cavity - the blastocoele and the ICM are thereafter called the yolk cavity and egg cylinder. As the ICM is pushing downwards the primitive endoderm around the epiblast begins to proliferate and eventually surrounds the egg cylinder (visceral endoderm) and lines the inner surface of the mural trophectoderm (parietal endoderm).

By 6.5 days the third germ layer, the mesoderm, makes its appearance. The first mesoderm cells form at the posterior end of the embryo, i.e. the primitive streak. They multiply rapidly wedging their way laterally between the ectoderm and endoderm towards the anterior end of the egg cylinder in a caudo-cranial fashion. Some cells also push in between the extraembryonic ectoderm and the adjacent endoderm thus leaving the embryo proper and mainly become involved in the formation of the yolk sac.

Once the three germ layers are established (see Fig. 1.2) organogenesis begins. The head process is formed at the anterior end of the primitive streak and gives rise to the notocord (the notocord is the axis about which the vertebral column is laid down) and part of the lining of the gut. The neural groove which gives rise to the central nervous forms dorsal to the notocord. Figure 1.3 gives a schematic representation of early post-implantation development.

The foregut and hindgut form as invaginations in the endoderm at both the anterior and posterior ends of the embryo. There is a progressive drawing together of the endodermal lining of the foregut and hindgut which contributes to the midgut. The invaginations of the foregut pushes the endoderm and overlying mesoderm and ectoderm as a bulge into the amniotic cavity. This bulge is the head fold. There is rapid growth of neural folds in this region, which will form the brain.The heart will develop from the mesoderm of the head fold.

Around 7.75 days the somites begin to form as paired segmental structures in the paraxial mesoderm a little anterior to the primitive streak and are the first indication of metamerism in the developing foetus. Initially pairs of somites are formed at a rate of one per hour then the rate formation slows to one every 2-3 hours (Tam, 1981) and by day 13 all the somites (65 pairs) have been formed; of these 30 will form the skeleton, musculature and dermis of the body and the remaining 35 will form these structures in the tail. By 10 days pc the major elements of the circulatory system have developed from the paraxial mesoderm, and when the embryo turns, thus causing the hindgut and foregut to join, the gut is completed. During the next 10 days of development the embryonic organ systems undergo progressive differentiation.

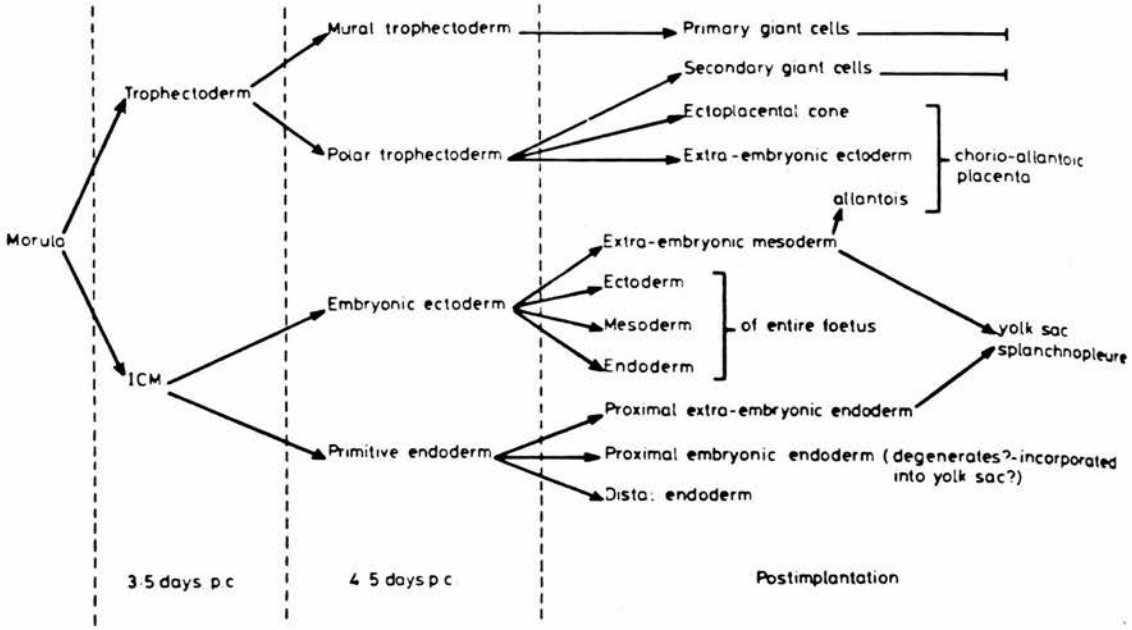


Figure 1.2. Principal cell lineages in the pre- and post-implantation mouse embryo (reproduced from Wilkins, 1986).

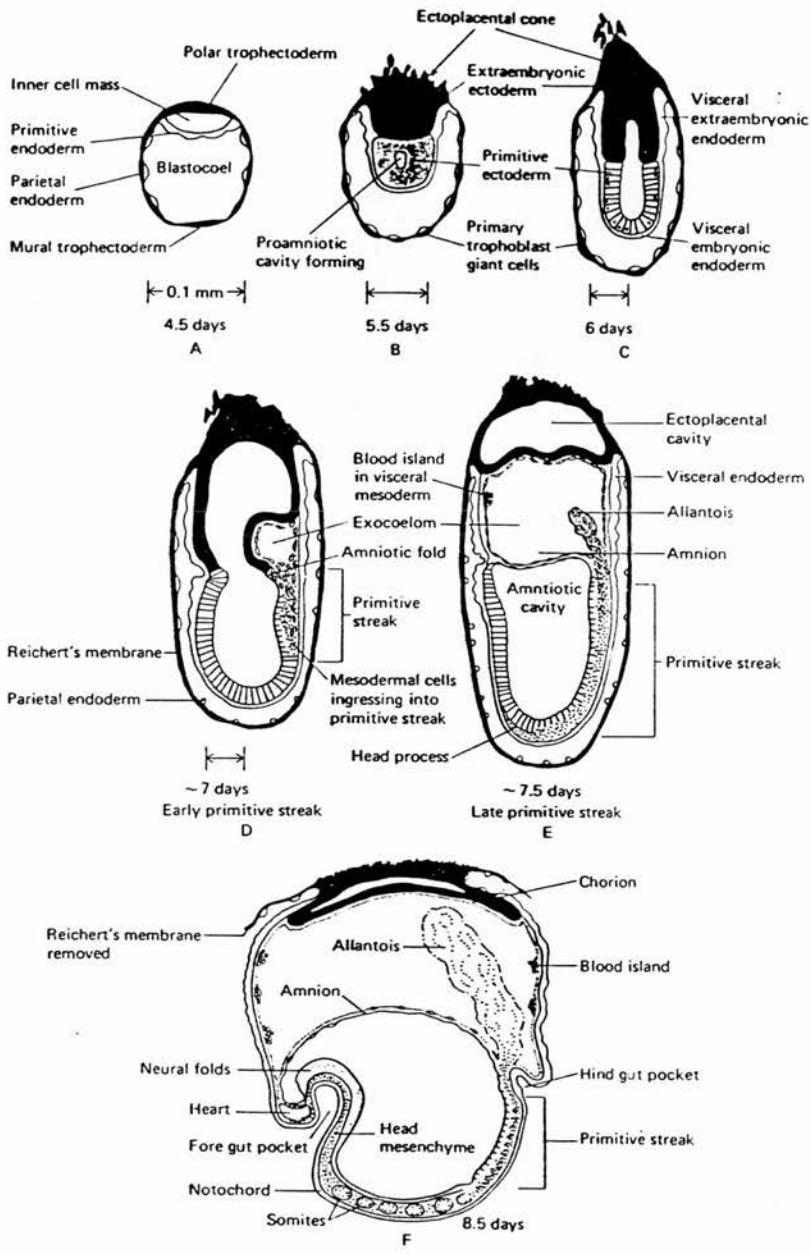


Figure 1.3. Schematic representation of early post implantation development of the mouse. The time of development is given in days *post coitum*. The timing is based on the C3H/He inbred line (reproduced from Hogan *et al.*, 1986).

1.1.1 *Gastrulation and Mesoderm Segmentation*

Gastrulation in the mouse begins at 6.5 days p.c with the appearance of the primitive streak which is located in the posterior third of the embryo and therefore defines the future A-P axis of the developing embryo. The primitive streak is where ectodermal cells invaginate to form the mesodermal layer which subsequently comes to lie between the ectodermal and endodermal cell layers via anterior and lateral migration. Those cells which migrate most laterally form the lateral plate mesoderm whereas those cells which lie closest to the primitive streak form the paraxial mesoderm. Cells which emerge from the anterior region of the primitive streak have an epithelial like structure and are distributed along the mid-line of the anterior part of the egg cylinder. This epithelial like structure is first apparent at 8 day p.c. and is known as the head process which is the precursor of the notochord. After formation of the notochord two important events occur. One is the formation of the neural folds which form from overlying ectoderm in response to an inductive signal from the underlying mesoderm (neural induction). The neural folds subsequently fuse to form the neural tube from which the CNS develops. The second important event to occur is the condensation of paraxial mesoderm on either side of the notochord to form the somites. Lateral to the somites, vesicles form in the lateral plate mesoderm which eventually coalesce to form the coelom. The mesoderm which is dorsal to the coelom and is associated with the ectoderm is called the somitic mesoderm whereas the mesoderm ventral to the coelom and therefore associated with the endoderm is called the splanchnic mesoderm (see Fig. 1.4)

The first somites appear at 7.75 days (Theiler, 1972). The remaining somites form in an anterior-posterior (cranio-caudal) sequence over the next few days (Flint, 1977). Somites form in the presomitic mesoderm which is posterior to the most

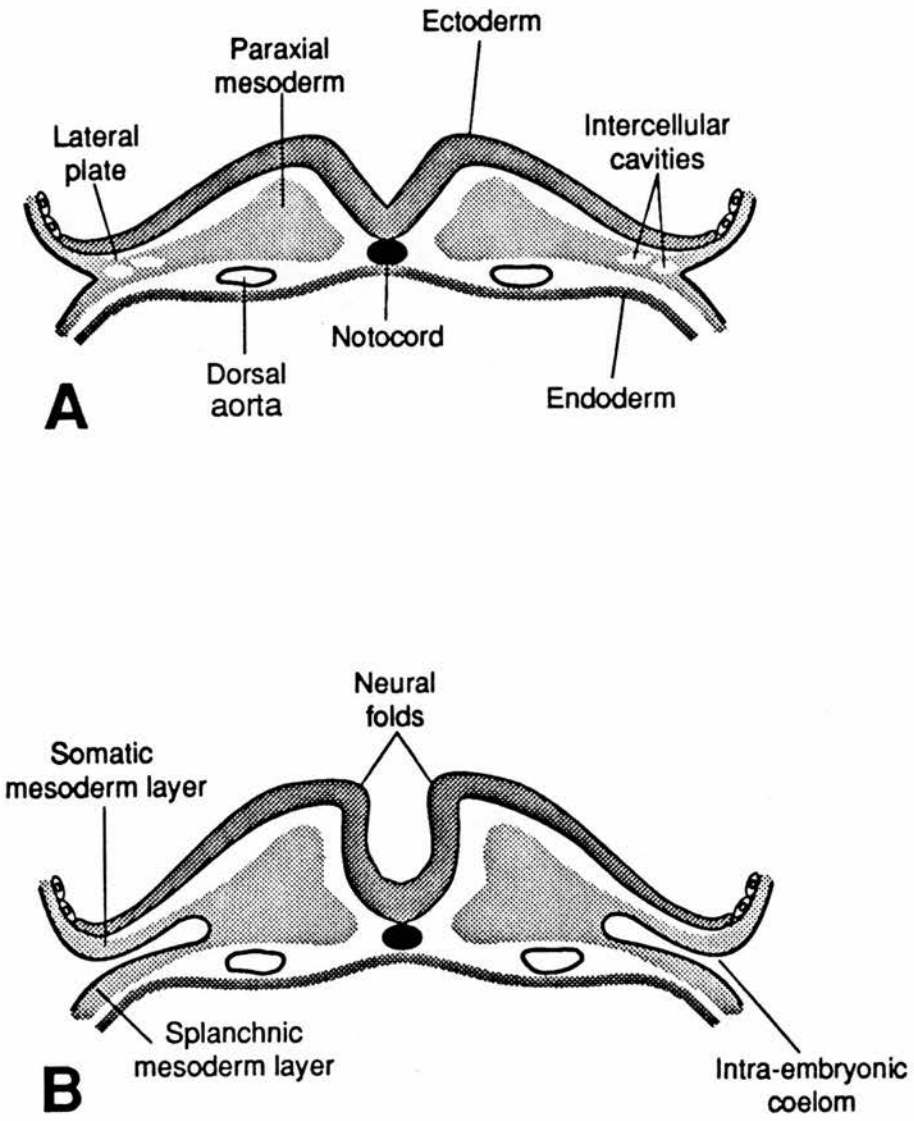


Figure 1.4. *A*, Transverse section through a mouse embryo. The vesicles of the developing coelom are visible in the lateral plate. *B*, The lateral plate mesoderm is now divided into the somatic and splanchnic mesoderm layers which line the intra-embryonic coelom.

recently formed somites and is contiguous with the posterior end of the embryo. It is thought that active recruitment of cells to the presomitic mesoderm occurs through the activity of the primitive streak and at a later stage via the activity of the tail bud (Flint *et al.*, 1978; Tam, 1981).

Although it was originally thought that the presomitic mesoderm was unsegmented it is now apparent that this is not the case. Scanning electron microscopy revealed the presence of whorls of cells, in the chick, which were subsequently called somitomeres (Meier, 1979). Since then somitomeres have been described in the head region and presomitic mesoderm of a variety of organisms (reviewed by Tam & Beddington, 1986) as well as the mouse (Meier & Tam, 1982; Tam *et al.*, 1982). The number of somitomeres seen in presomitic mesoderm correlates closely with the numbers of somites formed - if presomitic mesoderm is explanted and cultured the number of somites formed is usually the same as the number of somitomeres which had been detected (Tam & Beddington, 1986; Tam, 1986). Indeed the formation of somites is always associated with the disappearance of the corresponding somitomere in the presomitic mesoderm (Tam, 1986). Somite formation occurs in the absence of any signals from the rest of the embryo and, as somitomeres have been shown to exist in the earliest primitive streak stage mouse embryos (Tam & Meier, 1982), suggests the somite determination occurs early on in development. Once the somitomeres are established the future boundary and thus the position of the somite is determined. The number of cells from which the entire somite and all its derivatives will develop is also specified. It has however been demonstrated that the number of cells found in different somitomeres is not uniform. The newly segmented somites increase in size along the A-P axis. The lumbar and sacral somites are the biggest whereas the caudal somites get progressively smaller in size (Tam, 1981). This variation in size is due to a variation in the numbers of cells being incorporated into each somite at the time of segmentation

although how cell allocation is controlled is not yet known (Beddington & Tam, 1986).

Each mature fully condensed somite can be initially morphologically divided into an anterior and a posterior half (Meier, 1979 ; Tam *et al.*, 1982). Later on somites can be further subdivided into three regions: the sclerotome, the dermatome and the myotome. The sclerotome cells become arranged around the neural tube and notochord. The anterior sclerotomes from one pair of somites fuse with the posterior pair of the preceeding pair to form one vertebra (Balinsky, 1975). The vertebral column can be divided into five anterior to posterior regions - cervical (7), thoracic (13), lumbar (6), sacral (4) and caudal vertebrae (30-31). The thoracic vertebrae also give rise to the ribs. The myotome differentiates into the muscles of the body and limbs while the dermatome gives rise to the connective tissue of the body wall and the dermis. Thus as development proceeds the segmented nature of the paraxial mesoderm of the mouse is obscured and in the adult the only recognisible signs of segmentation lie in structure of the vertebral column and associated ribs and muscle.

Mentioned above was the formation of the neural folds, from which the CNS develops, via an inductive interaction between mesoderm and overlying ectoderm. To a large extent vertebrate embryos develop normally via inductive interactions between different cells and tissues (Jacobsen & Slater, 1988; Smith, 1988). The *Dh* mutation is classically regarded as a mutation the phenotype of which results from a failure of the splanchnic mesoderm to properly induce various tissues (Green, 1967; Russell, 1985). Therefore before going on to discuss the *Dh* mutation the next section reviews embryonic induction with particular reference to the formation and inductive properties of the mesoderm.

1.2 Embryonic Induction

The central question for developmental biologists is how the fertilised egg develops into an embryo with the correct spatial pattern of cellular differentiation. Development depends on cell interactions and most species use a mechanism involving a sequence of inductive interactions whereby new cell types are formed via interactions between existing cell types. Embryonic induction is defined as an interaction between one inducing tissue and another responding tissue as a result of which the responding tissue takes a course of differentiation it would not have followed had the interaction not occurred and has been described as "probably the single most important mechanism in vertebrate development." (Gurdon, 1987).

As mentioned previously *Dh* is classically regarded as a mutation, the phenotype of which, results from a failure of the splanchnic mesoderm to properly induce various tissues (Green, 1967; Russell, 1985). In vertebrates, as discussed below, the mesoderm has a major role in the control of pattern formation, particularly early in development when the body plan is being established. Thus in order to understand how *Dh* may act it is relevant to consider some of the cell interactions crucial to normal development, i.e. formation of and inductive properties of the mesoderm.

Embryonic induction was first described almost a century ago by Spemann (1901) and Lewis (1904) who both established that in a certain species of frog, lens formation from overlying ectoderm was dependent on the influence of the underlying optic vesicle of the embryonic brain. Over the years much work has been carried out on embryonic induction. Research has tended to focus on amphibia, particularly the frog *Xenopus laevis* as a model of choice for several reasons. Firstly, amphibian embryos are large therefore making experimental manipulations relatively easy. Secondly, the embryos are accessible to manipulation at all developmental stages and,

finally, as embryonic blastomeres survive on yolk reserves, they can divide and even differentiate in simple salt solutions. This allows one to test the effects of defined molecules on differentiation, without interference by serum components.

In normal amphibian development the basic body plan arises through the formation of three axes and a series of sequential inductive interactions (reviewed by Smith, 1988). The animal and vegetal poles represent the first signs of polarity in the egg. This is followed by the dorso-ventral polarity which is established shortly after fertilisation by a rotation of the eggs cytoplasm that is triggered and orientated by sperm entry point in the animal hemisphere (Gerhart *et al.*, 1984). The third axis, which imposes an anterior-posterior polarity on the developing embryo, occurs during gastrulation when the mesoderm, which initially lies in the middle of the blastula, migrates up and inside the gastrula finally lying immediately beneath the ectoderm. The development of the basic body plan is also dependent on a series of sequential inductive interactions throughout organogenesis (Jacobsen and Slater, 1988). A tissue which is induced at one developmental stage may itself become an inducer at a later developmental stage. Indeed cells of the the ectodermal lineage are subject to at least three well studied sequential inductions the first of which results in mesoderm formation (mesoderm induction). This is followed by neural induction which occurs at the mid-gastrula stage when mesodermal cells at the dorsal blastopore lip and inside the embryo induce overlying ectoderm to form neural tissue (Suzuki *et al.*, 1986; Kintner and Melton, 1987). The third well studied inductive interaction in amphibian development is optic induction which occurs when the optic lobes of the embryonic brain induce overlying ectoderm to invaginate and form the lens vesicles (Spemann, 1901; Lewis, 1904; Jacobson, 1966). Of these three inductive interactions particular attention has focused on mesoderm induction as this is one of the first inductive interactions in amphibia taking place within the first few hours of fertilisation and probably occurs in all vertebrates (Smith, 1988).

As mentioned above, the animal and vegetal poles represent the first signs of polarity in the amphibian egg. Fate mapping (Dale and Slack, 1987a) and isolation experiments (Nieuwkoop, 1969; Nakamura *et al.*, 1970) have shown that these poles have different developmental fates: the cells of the animal pole will form ectoderm whereas cells from the vegetal pole will form endoderm. Neither pole cultured in isolation will form mesodermal cells which constitute the third germ layer. However, when animal cells are placed in contact with vegetal pole cells mesodermal cells are formed from the equatorial animal pole cells (Nieuwkoop, 1969; Sudarwati & Nieuwkoop, 1971; Dale *et al.*, 1985) suggesting that animal pole cells have the ability to respond to but not to produce a mesoderm inducing signal and that vegetal cells are responsible for production of the inducing substance.

Mesoderm induction occurs at the mid-blastula stage, a time at which dorso-ventral polarity has been established. When dorsal and ventral mesoderm are cultured in isolation before gastrulation they exhibit quite different developmental fates. Dorsal mesoderm gives rise to notocord and muscle whereas ventral mesoderm gives rise to blood and mesenchyme (Boterenbrod & Nieuwkoop, 1973; Dale & Slack, 1987b). These results suggest that there are at least two distinct regions in the vegetal hemisphere which are transmitted to the animal hemisphere as result of mesoderm induction; i.e. dorso-vegetal cells induce dorsal-type mesoderm whereas ventro-vegetal cells induce ventral-type mesoderm.

However although large muscle blocks are characteristic of dorsal mesoderm induction the fate map of *Xenopus* shows that more than half the somite tissue comes from the ventral half in normal development (Dale and Slack, 1987a). This observation has prompted the idea that a second inductive interaction which has been termed "dorsalisation" is responsible for further sub-dividing the mesoderm (Slack, Dale & Smith, 1984; Smith *et al.*, 1985). Dorsalisation can be shown demonstrated by

combining dorsal and ventral marginal zones of gastrulae. In such combinations large muscle blocks are formed from the ventral tissue which, if cultured in isolation, differentiates mainly into mesenchyme and blood (Slack & Forman, 1980; Smith & Slack, 1983). This dorsalizing effect can be further demonstrated by grafting dorsal marginal zone into host ventral marginal zone and producing a double dorsal embryo with a secondary axis with much of the somitic mesoderm (normally a dorsal mesoderm derivative) being derived from host ventral marginal zone (Spemann & Mangold, 1924; Smith & Slack, 1983). If the converse experiment is performed a double ventral embryo is not formed. This reinforces the idea that the dorsal marginal zone is a special signal region that controls the developmental fate of adjacent ventral tissue (Smith & Slack, 1983). These experiments suggest that the formation of a normal set of axial mesodermal structures requires a dorsalising signal from a restricted sector of the blastula/gastrula dorsal marginal zone (the 'organiser') and have formed the basis for the three signal model of mesoderm induction (Dale & Slack, 1987b). The first two inductive signals are those of Boterenbrood and Nieuwkoop (1973); i.e., dorsal vegetal cells induce predominantly notochord with small amounts of muscle whereas the ventral cells induce blood, mesothelium and mesenchyme. The third signal originates from the newly induced dorsal mesoderm which dorsalises adjacent ventral mesoderm so that muscle rather than blood is formed.

The anterior-posterior (A-P) axis of the developing embryo is laid down during gastrulation when the mesodermal cells at the dorsal blastopore lip, a structure analogous to the avian or mammalian primitive streak (Hamburger, 1988), migrate up and inside the embryo so that they finally come to lie beneath the ectoderm. As the embryo gastrulates the A-P fates of the newly formed mesodermal cells are established. It has been shown that those mesodermal cells which migrate first, and therefore the furthest, will induce anterior / head structure. Cells which begin migration later, and therefore migrate the least, will induce posterior / tail structures (Mangold,

1933). However this fate is not absolute. If the normal gastrular migration is disrupted so that prospective anterior mesoderm stops midway these cells will differentiate into trunk structures and a headless embryo is formed (Gerhart *et al.*, 1984). However, as discussed below, determination of A-P fate is not simply a consequence of where the cell should happen to end up.

At the start of gastrulation transplantation of uninvaginated cells at the dorsal blastopore lip into the ventral marginal zone of an early gastrula host will result in the induction of anterior structures and an entire secondary axis (Spemann & Mangold, 1924). Later work by Smith and Slack (1983) established that these cells are actually specified as anterior / dorsal mesoderm. If cells from the blastopore lip are transplanted at later stages of gastrulation they are still able to induce and organise host mesoderm but with decreasing ability (Ruiz i Altaba & Melton, 1990) suggesting that the A-P fate of the mesoderm is, in part, determined as it invaginates. The A-P fate of mesoderm is therefore determined during gastrulation and depends on invagination and final position. However there is evidence that there is a third, maternal, component involved in the determination of A-P fates.

UV irradiation of a fertilised egg causes a delay in the onset of gastrulation and dose dependent anteriorly deficient embryos (Gerhart *et al.*, 1986). Treatment of the egg with lithium also yields a dose dependent effect but this time posteriorly deficient embryos are formed (Kao *et al.*, 1986). Both of these treatments would appear to be affecting maternal factors as they only produce the above effects when the eggs are treated at a stage when embryonic transcription has not yet begun.

1.2.1 *Timing and Localisation of the Response*

Inductive interactions occur throughout the development of the embryo and can be described as either permissive or instructive (Saxen, 1977). A permissive induction is one where the responding tissue is so commited to its final fate that any unspecific signal will result in successful induction. Instructive induction on the other hand occurs when the responding tissue is relatively uncommited and requires a specific signal. Thus in permissive inductions the response is primarily influenced by the responder whereas instructive inductions are more influenced by the inducing tissue even though the responding tissue may itself be limited in the ways it can differentiate (reviewed by Gurdon, 1987). It would therefore seem likely that instructive inductions are more prevalent early in development and as development proceeds and developmental fates become more limited permissive inductions play a larger role.

The timing and localisation of inductive interactions is of paramount importance in the development of a normal embryo. Although the exact time at which the ability to induce and respond is acquired is unclear it is known that both capabilities, in both early and late inductions, are lost shortly after the interaction would have occurred *in vivo* (Gurdon, 1988; Dale *et al.*, 1985; Rawles, 1963).

The localisation of any inductive response is obviously of great importance in order that only those cells which should contribute to the tissue / cell type do so. It is known that more cells have the ability to respond than actually do, for example all the blastula cells of the animal hemisphere are capable of forming mesoderm but only about 40% actually do so (Dale *et al.*, 1985). The mechanism by which the number and position of responding cells is limited is not clear although there are a few hints.

It is clear that proximity of the responding and inducing tissue is important rather than physical contact (Grunz & Tacke, 1986; Gurdon, 1989). It has been suggested

that the inducer reaches, by diffusion, adjacent responding cells beyond which its effect is further spread by diffusion and / or homoiogenetic induction. Homoiogenetic induction describes the phenomenon whereby cells induced to differentiate in one way can induce neighbouring cells to do the same (Mangold & Spemann,1927). Although homoiogenetic induction is usually thought to contain an amplification step one can envisage how the inductive response is localised if the inducing signal is passed on without amplification as the response would cease when the amount passed on went below a certain threshold. The localisation of the response is then further refined by a variety of factors including the slow diffusion and / or instability of the inducer, loss of competence of the responding cells and the amount of inducer released over a restricted time period.

Another mechanism thought to play a role in localising the inductive response is the community effect whereby the ability to respond to induction is enhanced or even dependent on neighbouring cells differentiating in the same way (Gurdon, 1988). This community effect would help ensure that any peripheral cells would contribute to either a responding group of neighbours or to the adjacent group all of which fail to respond.

Thus in conclusion it can be said that the formation of, and subsequent inductive properties of, the mesoderm play a major role in the establishment of the body plan. One can therefore envisage how, as is the case with *Dh* mice, defective mesoderm can disrupt the normal body plan. Induction is part of a continuum of developmental processes, the timing and localisation of which needs to be precisely co-ordinated with many other events in development. It is precisely these processes which are classically thought to be defective in *Dh* mice, i.e. malformation of the splanchnic mesoderm with the subsequent loss of co-ordinated inductive interactions. The next section reviews the classical view of the nature of the *Dh* phenotype and presents an updated hypothesis which suggests that, although failed inductive interactions play a major role in

producing the pleiotropic phenotype of *Dh*, the primary defect may occur during somitogenesis.

1.3 The *Dh* Mutation

The *Dh* mutation, which was first described in detail in 1959 by A.G.Searle, is a member of the luxoid group of mouse mutants which are characterised by a twisting of the fore- or hind-limbs and the reduction or loss of certain long bones (Gruneberg, 1963). Members of this group which includes *luxate* (Carter, 1951; 1953; 1954) and *luxoid* (Forsthoefel, 1958; 1959), also exhibit a wide variety of visceral abnormalities, the developmental relationship of which, when first described, were difficult to relate to the limb abnormalities.

In heterozygotes, *Dh* causes pre-axial abnormalities of the hind limb only, asplenia, a reduced number of pre-sacral vertebrae and ribs, a shortened coelom, a small and defective digestive tract and abnormalities of the urogenital system (Searle, 1964; Green, 1967). In homozygotes the abnormalities are more severe and are usually fatal. In 1967 Green demonstrated that in *Dh* /+ and *Dh* / *Dh* animals the earliest morphological defect was in the splanchnic mesoderm at 9.5 days. She postulated that the primary effect of the *Dh* locus was in the splanchnic mesoderm and that the pleiotropic effects of the *Dh* mutation were all traceable to this tissue. This hypothesis assumes that the *Dh* gene interferes with the normal structural arrangement of the cells of the splanchnic mesoderm.

At 9.5 days the region of the splanchnic mesoderm from the anterior end of the coelom to the posterior end of the primordium of the dorsal pancreas, referred to by Green (1967) as the anterior splanchnic mesodermal plate (ASMP), consists of a thick

columnar epithelium. In *Dh* /+ and *Dh* / *Dh* embryos the ASMP was found to be defective or missing.

In *Dh* /+ and *Dh* / *Dh* embryos the ASMP does not extend as far posteriorly as normal and is absent from the region where the spleen would normally develop. Green therefore concluded that the ASMP and development of the spleen were related although it was not known whether the epithelial cells of the ASMP were precursors or inducers of the spleen. Likewise Green postulated that a loss of normal inductive activity, due to abnormal structure / lack of epithelial organisation, of the splanchnic mesoderm resulted in the failure to promote the full growth of the gut endoderm. The disorganisation of the mesoderm cells was also thought to hinder the coelom from extending to its full posterior position.

It had previously been shown that *luxate* mice have a reduced number of pre-sacral vertebrate and also have a shortened coelom (Carter, 1954) whereas *luxoid* mice have an increased number of pre-sacral vertebrae and a longer coelom (Forsthoefel, 1959). Green thus concluded that the decreased number of vertebrae in *Dh* /+ and *Dh* / *Dh* mice was related to the the shortened length of the coelom. She postulated a sequence of events whereby the shortened coelom resulted in the anterior displacement of umbilical artery which she suggested may have been the actual hind limb bud inductor. Normally the coelom elongates posteriorly and Green postulated that this process normally exerts some pressure on the umbilical arteries. In mutant animals because the posterior end of the coelom is shortened there is a loss of pressure and consequently an anterior displacement of the umbilical arteries. Consequently the hind limb bud which bears a constant spatial relationship to the umbilical artery (Green, 1967; Rooze, 1977) also becomes displaced anteriorly and as a direct consequence so does the pelvic girdle. The reduced number of vertebrae were thought to have ocurred

as a direct failure of the correct inductive effect of the pelvic girdle on the axial skeleton.

Finally the differentiation of the ureter, bladder, urethra and vagina occurs in the mesoderm which surrounds the posterior end of the coelom. The urogenital abnormalities described by Searle (1964) which occur occasionally in heterozygotes and almost without exception in the homozygotes were thought to due either the abnormal / loss of normal inductive influence of the mesoderm surrounding the posterior end of the mesoderm or to the faulty spatial relationship caused by the short and abnormally shaped coelom.

Thus central to Green's hypothesis is the fact that in the mutant animals the splanchnic mesoderm is disorganised or missing and consequently normal inductive interactions fail to occur. However this hypothesis was written some twenty three years ago and obviously since then our knowledge of developmental biology has expanded. Although certain aspects of Green's original hypothesis may still hold true, in light of more recent work certain aspects seem dubious. For example it has been shown that the pre-somitic mesoderm is pre-determined with respect to the type of vertebrae if forms (Kieny *et al.*, 1972). It therefore seems unlikely that the pelvic girdle can have an influential inductive role on the axial skeleton. In view of more recent work it is possible to put forward not so much an alternative hypothesis as an updated one.

Green showed that, even in *Dh* /+ and *Dh* / *Dh* , the hind limb bud bears a constant spatial relationship to the umbilical artery, an observation later confirmed by Milaire (see Rooze, 1977). In +/+ mice the limb bud forms at the level of somites 23-28 (Theiler, 1972). In *Dh* /+ mice the hind limb bud is displaced anteriorly so that it comes to lie at somite level 20-27 whereas in *Dh* / *Dh* mice the hind limb buds are even further displaced and form at the level of somites 19-26 (Green,1967). The loss

of pre-sacral vertebrae had previously been shown by Searle (1964) to be more extensive in *Dh / Dh* than *Dh* /+ animals and was due to a tendency for the loss of lumbar and thoracic vertebrae.

Although lateral plate mesoderm is known to be regionalised since it gives rise to fore- and hind limb buds opposite specific somites, (Theiler, 1972), it was not known whether the segmentation seen in the paraxial mesoderm, i.e. formation of somitomeres and somites, extended laterally into the intermediate and lateral plate mesoderm. Meier (1980) has reported the presence of segmentation in the lateral plate mesoderm. He described the lateral plate mesoderm as having " a series of undulations with grooves" which pointed towards the intersegmental boundaries of the somitomeres. He also noted that these grooves reached from the somatic mesoderm across the primitive coelom and contacted the splanchnic mesoderm. Vascular elements in the intermediate mesoderm were shown to extend into both the intersomitic boundaries and the lateral plate mesoderm. The above observations suggest that although the paraxial and lateral plate mesoderm originate from different regions of the primitive streak (Tam & Beddington, 1987), there is some communication between them. One can therefore envisage that the loss of somites might have an effect on the lateral plate mesoderm such that the normal organised structure of the mesoderm is disrupted. This in turn may effect the posterior position of the coelom and thus the position of the umbilical artery.

As mentioned previously the position of the hind limb bud and the umbilical artery share a constant spatial relationship. Green postulated that the umbilical artery was the limb bud inductor and that limb abnormalities were the result of impaired inductive reaction. Hind limb buds emerge as thickenings in somatic tissue at somite levels 23-28 (Theiler, 1972). In the chick it has been shown that the somatic mesoderm requires the presence and stimulation of the adjacent somitic mesoderm to be

morphogenetically active (Pinot, 1970; Kieny, 1971). This work was confirmed in the mouse when Agnish and Kochhar (1977) showed that limb buds from 12 and 13 day embryos were capable of self - differentiation but a cellular contribution from the adjacent somitic mesoderm was necessary and critical for the proper growth and differentiation of limb buds from 11 day embryos. Thus if, as is the case of *Dh* mice, the limb bud is displaced this crucial early interaction would be impaired.

The above scenario postulates that the primary effect of the *Dh* gene occurs during somitogenesis, i.e. loss of somites, and that the disorganisation of the splanchnic mesoderm is a direct consequence of this. The pleiotropic effects of *Dh* occur due to an impairment / failure of later inductive events. The loss of pre-sacral vertebrae has previously been shown to be more extensive in *Dh/Dh* than *Dh* /+ animals (Searle,1964), although whether this was due to the loss of specific vertebrae was not examined. If, as the above suggests, *Dh* is involved in controlling somite number it would suggest that this gene, during normal development has a role analogous to the *Drosophila* homeobox-containing segmentation genes.

1.4 Homeobox - Containing Genes

Spatial and temporal organisation is of the utmost importance during development and when one considers the efficiency and reproducibility of development it would suggest that a precise and finely tuned set of control mechanisms are being put into play. Much of the work done on understanding early development has concentrated on an analysis of gene activity, both gene expression and control and has centred on *Drosophila melanogaster* . The fruitfly has been used extensively because the mammalian system is beset with problems, the most prevalent being the inacessability and complexity of the embryo. The study of developmental mutants in

Drosophila has revealed the presence of three major and distinct classes of developmentally important genes: (1) maternal effect genes; (2) segmentation genes; (3) homeotic genes (Lewis, 1978; Kaufman *et al.*, 1980; Nusslein-Volhard & Wieschaus, 1980).

It has been demonstrated that the segmented pattern of the *Drosophila* embryo is primarily organised via the activity of two centres localised to the anterior and posterior poles of the embryo (Frohnhofer *et al.*, 1986). Transplantation experiments have shown that these centres have inductive properties and can reorganise pattern and polarity when transplanted ectopically (Nusslein-Volhard *et al.*, 1987). A small set of maternally coded genes have been described which are required for these activities (reviewed by Nusslein-Volhard *et al.*, 1987). One such maternally encoded gene is *bicoid* (*bcd*) which is involved in anterior body pattern. The protein product of *bcd* is a graded morphogen with a peak of expression at the anterior end of the pole of the embryo (Frohnhofer & Nusslein-Volhard, 1986; 1987; Nusslein-Volhard *et al.*, 1987; Driever & Nusslein-Volhard, 1988a; 1988b). In embryos which are derived from *bcd*$^-$ females the head and thoraic structures are missing. Maternal effect genes have also been described which are involved in posterior development (e.g. *oskar*, Schupbach & Wieschaus, 1986) and both anterior and posterior development (e.g. *torso*, Schupbach & Wieschaus, 1986). These maternal genes provide only approximate information about the axes and general organisation of the embryo. Precise pattern formation occurs with the activation of zygotic genes. The zygotic genes which are involved in pattern formation can be grouped into two classes: the segmentation genes and the homeotic genes.

The segmentation genes are involved in determining the number and polarity of the body segments and can be divided into three phenotypic classes according to the pattern defect involved (Nusslein-Volhard & Wieschaus, 1980). (1) Gap genes:

mutations in these genes result in the deletion of up to eight adjacent segments. Examples of this class include *Kruppel* (*Kr*) and *hunchback* (*hb*) (Nusslein-Volhard & Wieschaus, 1980; Scott & Carroll, 1987). (2) Pair rule genes: mutants of this class of genes develop with only half the normal number of segments as alternate segments are deleted. Examples of this class include *even-skipped* (*eve*) (Nusslein-Volhard, Wieschaus 1980) and *fushi-tarazu* (*ftz*) (Hafen *et al.*, 1984) and (3) Segment polarity genes: mutants in this class retain the normal number of segments but in each segment a particular region is deleted and the remainder is present as a mirror image duplication. Examples of this class include *engrailed* (*en*) (Kornberg *et al.*, 1985) and *wingless* (*wg*) (Baker, 1987).

Thus the gap loci can be viewed as coarse dividers of the embryo, pair-rule genes act to divide the embryos into the segmental units and the segment polarity genes are involved in pattern formation within each segment.

The homeotic genes constitute the third developmentally important class of genes. The homeotic genes are involved in conferring segment identity and mutation in these genes result in one body segment being replaced by another (Lewis, 1978). The homeotic genes are found within two clusters : the genes of the Antennapedia complex (ANT-C) control head and thoracic segments (Kaufman *et al.*, 1980) whereas those of the Bithorax complex (BX-C) control thoracic and posterior segments (Lewis, 1978).

One major feature of these developmentally important genes is the presence in several of them of a conserved region called the homeobox (McGinnis *et al.*, 1984). The homeobox is 180 nucleotides long and was first discovered in the *ftz* , *Antp* and *Ubx* genes (Kuroiwa *et al.*, 1984; Scott & Wiener, 1984; Gehring, 1985; Regulski *et al.*, 1985). The homeobox has since been described in various homeotic genes (Harding *et al.*, 1985), segmentation genes (Bopp *et al.*, 1986; Cote *et al.*, 1987) and maternal effect genes (Frigerio *et al.*, 1986; MacDonald & Struhl,1986). No

homeobox containing genes have been described for any of the gap genes although several of them have protein domains similar to the 'zinc-finger' DNA binding domains of *Xenopus* transcription factor IIIA (Rosenberg *et al.*, 1986; Tautz *et al.*, 1987). The prototype homeobox is defined by that found in the *antennapedia* gene (*Antp*), however divergent homeobox domains can be found in the *engrailed* and *invected* genes (Poole *et al.*, 1985) and the *paired* (*prd*) and *bcd* genes (Frigerio *et al.*, 1986). Furthermore some segmentation genes in addition to the *prd* type homeobox have a second conserved domain called the paired box (Bopp *et al.*, 1986) the function of which is unknown.

1.4.1 *The Homeobox Encodes a Transcription Factor*

Many homeodomain proteins have been found to be localised within the nucleus (White & Wilcox, 1984; Beachy *et al.*, 1985; Carroll & Scott, 1985; DiNardo *et al.*, 1985; Hoey & Levine, 1988) and there have been several reports of homeodomain proteins having DNA binding activity. *Engrailed* has been shown to bind to an upstream region of itself and *ftz* (Desplan *et al.*, 1985) and *Ubx* binds tightly to its own promoter as well as well as that of *Antp*, a gene which it is known to repress (Beachy *et al.*, 1988). Their are several other examples of DNA binding activity of homeodomain proteins (Struhl *et al.*, 1989; Laughon *et al.*, 1988; Hiromi & Gehring, 1987; MacDonald *et al.*, 1986).

The homeobox encodes the homeodomain a region which has shown to have homology with prokaryotic (*cro* and repressor proteins of lambda bacteriophage) and eukaryotic (yeast MAT a1 and MAT α2 proteins) genes (Shepherd *et al.*, 1984; Laughon & Scott, 1984). These genes encode DNA binding proteins therefore suggesting that the protein products of the homeobox genes exert their effect via DNA

binding, an argument which has been strengthened by the recent finding that the *Antp* and *engrailed* homeodomains consists of a helix-turn-helix motif (HTH) (Otting *et al.*, 1988; Qian *et al.*, 1989; Kissinger *et al.*, 1990). This motif is also present in the proteins described above and is responsible for their DNA binding ability (reviewed by Brennan & Mattews, 1989). However there are some major differences between the homeodomain and prokaryotic proteins with regard to the way in which they bind DNA.

The lambda repressor protein makes sequence specific contact with the major groove of DNA via the second helix (recognition helix) while the remaining helix holds it in place (Pabo & Sauer, 1984). The *engrailed* homeodomain on the other hand has three helices, which are longer than their lambda counterparts, and an extended N-terminal arm (Kissinger *et al.*, 1990). The N-terminal helix (helix 1) does not contact the DNA but acts as a stabiliser by packing against the second and third helices. Helix 2 contacts the sugar phosphate backbone whereas helix three makes sequence specific contact with bases in the major groove as well as several contacts with the sugar phosphate backbone. A particularly unique feature of the homeodomain is the contact with bases of minor groove DNA by a region upstream of helix 1 (Otting *et al.*, 1990; Kissinger *et al.*, 1990). If these N-terminal regions are deleted the binding affininty for the target DNA is greatly reduced (Percival-Smith *et al.*, 1990). As the homeodomain binds to both major and minor groove DNA it may not be able to bind to DNA within nucleosomes therefore increasing binding specificity (Kissinger *et al.*, 1990)

This DNA binding activity is mediated via the binding of the homeodomain to specific sequences. Using deletion constructs Desplan *et al.* (1988) were able to localise the DNA binding activity of *en* to seventy amino acids. This region begins at eleven amino acids N terminal to the homeobox and extends through the first fifty-nine

amino acids of the homeobox. The binding site consensus sequence TCAATTAAAT is found in clusters in the *en* regulatory region and binding is influenced by the primary sequence and the number of copies of the sequence. The *ftz* homeodomain also binds to this sequence but differs in its preference for related sequences. A different consensus sequence (TCTAATCCC) has been described for *bcd* binding activity (Driever +Nusslein-Volhard, 1989). This sequence is found in the upstream region of *hb*, a segmentation gene, the transcription of which, *bcd* controls (Lehmann & Nusslein-Volhard, 1987; Tautz, 1988; Driever & Nusslein-Volhard, 1989). A third consensus sequence (TAA)n has been described, to which *Ubx* binds (Beachy *et al.*, 1988).

Further evidence that the homeodomain is responsible for DNA binding and subsequent transcriptional regulation has been stengthened by three lines of evidence. Firstly substitution of residue nine of the recognition helix by two independent groups has shown that this particular residue is important in determining binding specificity. Hanes & Brent (1989) were able to switch the binding specificity of the *bcd* protein from *hb* to sites normally bound by *Ubx* and *ftz* homeodomain proteins. Similarly Triesman *et al.* (1989) changed the binding specificity of the *prd* homeodomain to that of *bcd* / *ftz* . Secondly Kuziora & McGinnis (1989) constructed a chimeric protein by replacing the *deformed* (*Dfd*) homeobox with that of *Ubx* . Normally these two genes positively self-regulate their own expression (Bienz & Tremml, 1988; Kuziora & McGinnis, 1989). The chimeric protein was unable to activate transcription of the *Dfd* gene but was able to activate *Antp* transcription, a gene normally repressed by *Ubx* (Boulet & Scott, 1988; Hafen *et al.*, 1984a). It has been proposed that the activation of *Antp* was a due to a region rich in acidic amino acids which is found in the *Dfd* protein upstream from the homeobox (Regulski *et al.*, 1987). Many proteins which activate transcription have two domains, one which binds to specific DNA sequences in the promoter of the target gene and the second which acts as a general transcription activator (Ptashne,1988). *Ubx* normally lacks this acidic general activator region

(Weinzerl *et al.*, 1987; Kornfeld *et al.*, 1989). That the *Dfd* portion of the chimeric protein can reverse the normal regulatory action of *Ubx* on *Antp*, i.e. change repression to activation, suggests that although the homeodomain is important for specific DNA binding flanking regions also play a role in transcriptional regulation. Homeodomain proteins can act as either activators or repressors of transcription (Krasnow *et al.*, 1989; Winslow *et al.*, 1989; Biggin & Tijian, 1989) and in some cases the transcriptional effect depends on which promoter they are binding (Krasnow *et al.*, 1989). Finally it has been shown that rat and human transcription factors contain divergent homeodomains (Ko *et al.*, 1988; Ingraham *et al.*, 1988; Bodner *et al.*, 1988; Scheiderеit *et al.*, 1988).

1.4.2 *Evolutionary Conservation of Homeobox Containing Genes.*

Using *Drosophila* homeobox probes cross hybridising sequences have been detected and in several cases shown to be developmentally expressed in a wide variety of organisms including *Xenopus,* mouse, rat and human (Levine *et al.*, 1984; McGinnis *et al.*, 1984). The vertebrate and *Drosophila* sequences can exihibt as much as 70-80% sequence conservation at the nucleotide level and this can increase at the amino acid level. The homeobox therefore seems to encode a domain which has been extensively conserved over a vast evolutionary span. These findings are very exciting as they suggest that at least some of the genes involved in controlling mammalian development are within reach.

A number of murine homeobox genes with homeoboxes similar to the *Drosophila* prototype defined by *Antp* have now been identified and these reside in four clusters each of which spans more than 100 kilobases - *Hox 1, Hox 2, Hox 3* and *Hox 4* clusters are on chromosomes 6, 11, 15 and 2 respectively (Duboule &

Dolle,1989; Duboule *et al.*, 1990). DNA sequence analysis has revealed that almost identical proteins are encoded in different clusters (e.g. *Hox 1.1* and *Hox 2.3*; Meijlinck *et al.*, 1987; Kessel *et al.*, 1987). Similar proteins can be grouped together as paralogous genes although not all clusters contain each paralogue. Furthermore the Antennapedia and Bithorax clusters of *Drosophila* can be aligned with the murine genes on the basis of homeobox sequence (Graham *et al.*, 1989; Duboule & Dolle, 1989).

One important feature of the *Drosophila* clusters is the relationship between the linear order of genes within the cluster and their developmental expression along the A-P axis (Harding *et al.*, 1985; Akam, 1987). The domains of expression along the A-P axis directly correspond with the segments affected by mutations in this genes (Harding *et al.*, 1985). Likewise the linear order of the murine genes within the *Hox* clusters has been shown to reflect its domain of expression, with respect to its anterior limit, along the A-P axis (Graham *et al.*, 1989; Gaunt *et al.*, 1988; Duboule & Dolle, 1989).

As well as these clusters there are several other homeobox genes in the mouse. These homeobox genes have more divergent homeoboxes than those found in the *Hox* clusters and include two *engrailed* related genes (Joyner & Martin, 1987; Hill *et al.*, 1987).

Evolutionary conservation of the homeobox genes would suggest a common origin and that they have a common function. This has quite recently been dramatically demonstrated via the expression, in *Drosophila* , of two *Hox* genes the proteins of which induced a developmental response characteristic of the *Drosophila* homologue (e.g. *Hox 2.2* transformations induced thoracic legs in place of antennae, Malicki *et al.*, 1990; McGinnis *et al.*, 1990). These experiments show that the structural similarities between the homeoboxes of *Drosophila* and vertebrate do translate into

similar regulatory specificities but it remains to be determined as to whether regulatory interactions and target genes are also conserved.

The vast amounts of data being accumulated have shown that these murine homeobox genes are expressed in a temporal and spatially restricted manner during development (reviewed by Holland & Hogan,1988; Kessel & Gruss,1990). One of the major setbacks with regard to the functional analysis of the homeobox genes in the mouse, has been the lack of associated developmental mutations which were so vital in establishing the function of the *Drosophila* genes (see Nusslein-Volhard & Wieschaus, 1980). However one intriguing and provocative outcome from mapping studies of the *En-1* and *En-2* genes was their close genetic linkage with known developmental mutations: *En-1* was shown to map close to or at the *Dh* locus (Hill *et al.*, 1987; Joyner & Martin, 1987) whereas *En-2* mapped close to a pair of genes - *Hemimelic extra toes* (*Hx*) and *Hammertoe* (*Hm*) Joyner & Martin,1987). This finding was all the more exciting as these three mutations exhibit similar skeletal defects (Searle, 1964; Knudsen & Kochhar, 1981; Green, 1989). This close genetic linkage raised the question of whether these genes were allelic and therefore *Dh* and *Hm/Hx* represented mutant alleles of their respective *En* gene

1.4.3 *Engrailed : Role in Drosophila and Mouse*

The *Drosophila* embryo is quite clearly segmented and each segment is comprised of an anterior and a posterior compartment. These compartments cannot be delineated anatomically like the segments but they can be identified via cell lineage studies. Cells of the different compartments are developmentally distinct and never mix (Garcia-Bellido et al, 1973; Lawrence *et al.*, 1979). Their importance as developmental

units has been shown by mutations, of which *engrailed* (*en*) is one, which disrupt development of only one of the compartments (Morata & Lawrence, 1977).

Engrailed (*en*) was one of the first segmentation loci to be identified in *Drosophila* (Eker, 1929). Genetic studies using chemical mutagens and X-rays have generated a large number of point mutants, breakpoint, deletion and insertion alleles (Nusslein-Volhard, 1980; Kornberg *et al.*, 1981a; Eberlein & Russell, 1983; Gubb, 1985). Most of the mutant alleles are fully recessive and the flies have a severely disrupted segmentation pattern which is often lethal (Kornberg *et al.*, 1981a). In these embryos pairs or large groups of segments fuse together and all the body segments are affected. The abnormality is in the posterior compartment: adult flies which carry non-lethal alleles or are gentically mosaic also exhibit abnormal segmental development but the abnormalities are confined to the posterior compartment. No *en* related defect has ever been detected in cells of the anterior compartment which develop normally in *en* mutants (Lawrence & Morata, 1976; Kornberg ., 1981a, 1981b; Lawrence & Struhl, 1982).

It has also been shown that the *en* mutation increases the size and shape of the posterior compartment and therefore fails to maintain the anterior - posterior compartment boundaries. The above observations suggest that *en* functions to establish the identity of the posterior cells and to maintain the anterior - posterior compartment boundary. *In situ* hybridisation which gives a very characteristic zebra like pattern (Ingham *et al.*, 1985; Kornberg *et al.*, 1985; Fjose *et al.*, 1985) and antibody localisation of the *en* protein (DiNardo *et al.*, 1985) have shown that *en* is only expressed in the cells of the posterior compartment.

As well as its role in segmentation and establishing posterior compartment identity wild type *en* function is also essential for the formation of the cellular blastoderm (Karr *et al.*, 1985). During the pre-cellular stage wild type *Drosophila*

embryos produce somatic, germ cell and yolk nuclei. In *en* embryos both the number and distribution of these nuclei is impaired. *In situ* hybridisation has failed to localise *en* expression to particular sections of the pre-cellular embryo (Weir & Kornberg, 1985). How this early *engrailed* function relates to its later position specific function is unclear.

Engrailed is known to encompass a functional genetic unit of at least 70 kilobases (kb) (Kornberg *et al.*, 1985; Kuner *et al.*, 1985) within which there is a small transcription unit of about 4 kb (Drees *et al.*, 1987). *Engrailed* rearrangments have defined lethal and non-lethal regions of 50 kb and 20 kb respectively (Kuner *et al.*, 1985).

Like many segmentation genes, *en* contains a homeobox, but the *en* homeobox is quite divergent from those found in the Antennapedia and Bithorax complexes (Poole *et al.*, 1985). The other striking feature of the *en* homeobox is the presence of an intron within the homeobox (Poole *et al* ,1985). An *en* related gene called *invected* (*inv*) is located just proximal to *en* (Poole *et al.*, 1985). These two genes share striking homology and their homeoboxes are almost identical - of eight amino acid differences five are conservative substitutions (Poole *et al.*, 1985; Coleman *et al.*, 1987). The homology between the two genes also extends beyond the homeobox in either direction (Coleman *et al.*, 1987). The 5' homology consists of 17 amino acids that are identical in both genes. This region is separated by sequences encoding 5 (*en*) or 26 (*inv*) nonconserved amino acids following which there is the near identical homeobox beyond which there is a stretch of 31 amino acids of which 26 are identical and the remaining 5 constitute conservative substitutions (see Fig. 1.5). *Inv,* like *en,* has an intron within the homeobox although the introns are not homologous (Poole *et al.*, 1985). *Inv* has the same developmental expression as *en* but its developmental role is not yet known (Coleman *et al.*, 1987).

b

```
                                                                  4
En-1                                                Pro Leu Val Trp Pro Ala
En-2                                                 *  Met Leu  *   *   *
en                                                  Asn Glu Met  *   *   *
inv                                                  *  Ile  *   *   *   *

                                                                      20
En-1   Trp Val Tyr Cys Thr Arg Tyr Ser Asp Arg Pro Ser Ser Gly  -
En-2    *   *   *   *   *   *   *   *   *   *   *   *   *   *   -
en      *   *   *   *   *   *   *   *   *   *   *   *   *   *   -
inv     *   *   *   *   *   *   *   *   *   *   *   *   *   *  Arg

            21                                  29

En-1    -  Pro Arg Thr Arg Lys Leu Lys Lys Lys Lys Asn Glu  -  Lys
En-2    -   *   *  Ser  *   *  Pro  *   *   *  Asn Pro Asn  -   *
en      -   *   *  Tyr  *  Arg Pro  *  Gln Pro  *  Asp Lys Thr Asn
inv    Ser  *   *  Ala  *   *  Pro  *   *  Pro Ala Thr Ser [ ] Pro

En-1   Glu Asp Lys Arg Pro Arg Thr Ala Phe Thr Ala Glu Gln Leu Gln
En-2    *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
en     Asp Glu  *   *   *   *   *   *   *  Ser Ser  *   *   *  Ala
inv     *   *   *   *   *   *   *   *   *  Ser Gly Thr  *   *  Ala

En-1   Arg Leu Lys Ala Glu Phe Gln Ala Asn Arg Tyr Ile Thr Glu Gln
En-2    *   *   *   *   *   *   *  Thr  *   *   *  Leu  *   *   *
en      *   *   *  Arg  *   *  Asn Glu  *   *   *  Leu  *   *  Arg
inv     *   *   *  His  *   *  Asn Glu  *   *   *  Leu  *   *  Lys

En-1   Arg Arg Gln Thr Leu Ala Gln Glu Leu Ser Leu Asn Glu Ser Gln
En-2    *   *   *  Ser  *   *   *   *   *   *   *   *   *   *   *
en      *   *   *  Gln  *  Ser Ser  *   *  Gly  *   *   *  Ala  *
inv     *   *   *  Gln  *  Ser Gly  *   *  Gly  *   *   *  Ala  *

En-1   Ile Lys Ile Trp Phe Gln Asn Lys Arg Ala Lys Ile Lys Lys Ala
En-2    *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
en      *   *   *   *   *   *   *   *   *   *   *   *   *   *  Ser
inv     *   *   *   *   *   *   *   *   *   *   *  Leu  *   *  Ser

        94
En-1   Thr Gly Ile Lys Asn Gly Leu Ala Leu His Leu Met Ala Gln Gly
En-2    *   *  Asn  *   *  Thr  *   *  Val  *   *   *   *   *   *
en      *   *  Ser  *   *  Pro  *   *   *  Gln  *   *   *   *   *
inv    Ser  *  Thr  *   *  Pro  *   *   *  Gln  *   *   *   *   *

                       114
En-1   Leu Tyr Asn His Ser Thr Thr Thr Val Gln Asp Lys Asp Glu Ser
En-2    *   *   *   *   *   *   *  Ala Lys Glu Gly  *  Ser Asp  *
en      *   *   *   *  Thr  *  Val Pro Leu Thr Lys Glu Glu  *  Glu
inv     *   *   *   *   *   *  Ile Pro Leu Thr Arg Glu Glu  *  Glu

        124
En-1   Glu ter
En-2    *   *
en     Leu Glu Met Arg Met Asn Gly Gln Ile Pro ter
inv    Leu Gln Glu Leu Gln Glu Ala Ala Ser Ala Arg ..... ter
```

Figure 1.5. Amino acid sequence analysis of the conserved region in *En-1*, *En-2* and the Drosophila genes *en* and *inv*. The sequences are aligned to achieve the best match. To accomplish this a space, denoted by a dash, was left between amino acids (aa) that are actually contiguous in *En-1* and *En-2*. In addition,22 aa in the *inv* sequence, denoted by brackets, are not shown. The dots represent 6 aa in the *inv* sequence immediately upstream from the stop codon. The homeobox is delineated by a box. An asterix denotes those aa in the *En-2*, *en* and *inv* sequences that are identical to the aa in the corresponding position in the *En-1* sequence. (Reproduced from Joyner & Martin, 1987).

Using *Drosophila en* probes analogous sequences have been described in the mouse (Joyner *et al.*, 1985; Joyner & Martin, 1987; Hill *et al.*, 1987). There are two *en* like genes in the mouse, *En-1* and *En-2* , but unlike *en* and *inv*, they are not linked: *En-1* maps to the central portion of chromosome 1 (Hill *et al.*, 1987; Joyner + Martin, 1987) whereas *En-2* maps to the proximal portion of chromosome 5 (Joyner & Martin, 1987).

En-1 and *En-2* exhibit a remarkable degree of homology to each other (79% identical at the nucleotide level) and to their *Drosophila* counterparts (Joyner & Martin, 1987). With regard to the homeobox there is 93% identity at the amino acid level between the two mouse genes and 70% identity when comparing all four genes. The homology between all four genes, including the centrally located homeobox, encodes 107 amino acids. Within this homologous region 78 (73%) of amino acids are identical in the mouse and *Drosophila* genes (see Fig. 1.5).

Both murine genes have been found to be expressed during embryogenesis. Using teratocarcinoma cells, which serve as an *in vitro* model for the mouse embryo at the peri-implantation stages (3.5 - 10.5 days of gestation, Martin, 1980), Joyner & Martin (1987) were able to show that both genes are expressed early in development. Using whole mouse embryo RNA, the transcripts were shown to be expressed throughout embryogenesis (up to 17.5 d) with a peak of expression between 10.5d and 12.5d. Transcripts of both genes were found to be abundant in 12.5d RNA from the posterior portion of the brain. These findings suggested that there were two features of *En* expression in the mouse that were similar to *en* expression in *Drosophila*. Firstly the expression of either gene is not exclusive (Coleman *et al.*, 1987) and secondly the genes may have a role in the development of the CNS - both *en* and *inv* have been detected in the CNS of *Drosophila* (Ingham *et al.*, 1985, Weir & Kornberg, 1981; Coleman *et al.*, 1987).

In situ hybridisation studies of mouse embryos (Davidson *et al.*, 1988; Davis *et al.*, 1988; Davis & Joyner, 1988) have confirmed that although *En-1* has a more diverse expression pattern than *En-2*, expression of either gene is not exclusive, as *En-2* seem to be expressed in a subset of *En-1* expressing cells (Davidson *et al.*, 1988). The diverse expression of *En-1* compared to *En-2* suggests that these genes are under differential control. Furthermore in those tissues where expression occurs, both genes are only expressed in restricted areas, an observation which suggests that cell position as well as cell type is an important regulatory factor (Davidson *et al.*, 1988). In several cases this restricted expression is within morphologically undifferentiated tissue suggesting that *En-1* and *En-2* may be involved in establishing and / or maintaining specific domains within developing tissues.

During development the expression of both genes begins at 8 days (d) in overlapping domains in the anterior neural folds - a period of early somite formation (Davis & Joyner, 1988). By 12d although both genes are expressed in similar regions of the developing brain *En-1* is also transiently expressed in the spinal cord, restricted region of the facial mesenchyme, limb buds, dermatome of the tail and in the sclerotome derived loosely packed cells of the pericordal tube which later gives rise to the vertebrae (Davis & Joyner, 1988; Davidson *et al.*, 1988). Interestingly the periodic expression of *En-1* throughout the length of the pericordal tube appears to be the exact complement of *Pax-1*, a mouse paired box containing gene, which is expressed in cells which eventually form the intervertebral discs (Deutsch *et al.*, 1988). It has been postulated that *En-1*, along with *Pax-1*, may be involved in vertebrate assembly (Davis & Joyner, 1988).

As mentioned above, both genes are expressed in the CNS. At 12d both genes are expressed in the developing midbrain and its junction with the hindbrain (Davidson *et al.*, 1988; Davis & Joyner 1988). Although *En-1* is expressed in other tissues, by

15.5d expression of both genes is primarily limited to the midbrain / hindbrain in overlapping but non-identical sets of differentiated cells (Davis & Joyner, 1988). In the adult mouse both genes are expressed in the same sets of cells in the pons but *En-2* is also expressed in the cerebellum (Davis & Joyner, 1988).

1.4.4 *Homeobox Genes and Induction*

Recent studies have indicated that homeobox-containing genes may constitute the first genetic response to induction. Inductive interactions are thought to be mediated by peptide growth factors (PGF) Thus far, the candidates fall into two classes.The first class consists of members related to transforming growth factor β and includes XTC-MIF which was isolated from a *Xenopus* cell line (Smith, 1987 ; Smith *et al.*, 1988 ; Rosa *et al.*, 1988), TGF-β2 (Rosa *et al.*, 1988) and perhaps the protein encoded by the maternal mRNA Vg1 which is localised in the vegetal hemisphere (Rebagliatti *et al.*, 1985 ; Weeks & Melton, 1987). Another member of this family, TGF-β1, has been shown to have no mesoderm inducing activity (MIF) itself (Slack *et al.*, 1987 ; Rosa *et al.*, 1988) but acts synergistically with members of the second class of MIFs - the fibroblast growth factor (FGF) family (Kimmelman & Kirschner, 1987; Rosa *et al.*, 1988). Active members of this family include both acidic and basic FGF (Slack *et al.*, 1987 ; Kimmelman & Kirschner, 1987).

More recently it has been shown that XTC-MIF has homology with mammalian activin A (Smith, 1990) which is known to induce mesoderm (Asashima *et al.*, 1990; van den Eijinden-Van Raaij *et al.*, 1990). Activins occur either as activin A ($\beta_A\beta_A$) or AB ($\beta_A\beta_B$); activin B has not been isolated from natural sources but synthetic activin B has properties similar to activin A (Mason *et al.*, 1989). Thomsen *et al.* (1990) cloned the *Xenopus* activin β_A and β_B chains and found that activin β_B transcripts were detected

in the *Xenopus* blastula whereas β_A transcripts did not appear until the late gastrula stage. They also found that recombinant *Xenopus* activin β_B protein induced mesodermal and neural tissue similar to that seen with both porcine activin A and murine phagocyte inhibitory factor (PIF) both of which have similar inductive / antigenic properties (Thomsen *et al.*, 1990; Sokol *et al.*, 1990). These results suggest that activin B rather than activin A plays a role in early induction and axial mesoderm patterning in *Xenopus*. Simultaneously the chicken β_B was also shown to be transcribed during a time of axial patterning (Mitrani *et al.*, 1990).

Although there is still a great amount of work required before we will know whether the above factors or their embryonic homologues are the true inducing substances one intriguing aspect of their inductive capacity which has come to light has been the demonstration that different PGFs induce mesoderm from isolated animal caps (i.e. the most distal region of the animal hemisphere) with different dorsal-ventral character. XTC-MIF induces more dorsal mesoderm than does bFGF and, in the case of XTC-MIF, the concentration used affects the amount and dorsal-ventral type of tissue induced (Smith *et al.*, 1988).

Recent studies have shown that in isolated *Xenopus* animal caps an early response to mesoderm induction by PGFs is the transcriptional activation of at least two homeobox genes : *Mix.1* (Rosa , 1989) and *Xhox 3* (Ruiz i Altaba & Melton, 1989a,1989b). The *Mix.1* gene is activated by XTC-MIF although its role in axis formation is not yet known. *Xhox3* on the other hand is known to be expressed in a graded fashion along the anterior-posterior axis in the axial mesoderm (Ruiz i Altaba & Melton, 1989a, 1989b). This graded expression is such that higher levels are in the posterior end and the lowest levels are found in the anterior end. This finding led to speculation about its role in A-P patterning. If large amounts of synthetic *Xhox3* mRNA are injected into fertilised eggs, embryos with anterior deficiencies are formed

indicating that the level of *Xhox3* RNA corresponds to the A-P character of the mesoderm (Ruiz i Altaba & Melton, 1989c). It is important to note that the above experiment did not disrupt the normal migration of mesodermal cells and one can therefore conclude that high levels of *Xhox3* mRNA inhibit the normal development of correctly positioned anterior cells. It is also worth noting that the above does not correspond to a homeotic mutation as the anterior cells are not transformed into posterior cells; i.e., low levels of *Xhox3* are apparently necessary for correct development of anterior structures but high levels are not sufficient to specify posterior development.

If homeobox containing gene expression is important in determining A-P fate then this begs the question of how this expression is being regulated. It has recently been shown that different PGFs induce different levels of *Xhox3* RNA (Ruiz i Altaba & Melton,1989b). XTC-MIF weakly induces *Xhox3* mRNA whereas bFGF induced 5-10x higher levels of mRNA. This differential induction of mRNA by XTC-MIF and bFGF is similar to the endogenous difference in *Xhox3* mRNA levels between anterior and posterior mesoderm of the early neurula embryo. As high / low levels of *Xhox3* expression are correlated to posterior / anterior determination respectively suggests that bFGF induces posterior and XTC-MIF induces anterior mesoderm. That homeobox genes are differentially activated has been further demonstrated by Cho & De Robertis (1990) who found that *Xhox1*, which is expressed in the anterior trunk region of the *Xenopus* embryo, is preferentially activated by XTC-MIF whereas *Xhox6*, a posteriorly expressed gene, is preferentially activated by bFGF. These findings further suggest that XTC-MIF has a role in anterior determination whereas bFGF plays a role in posterior determination.

Inductive interactions would therefore seem to involve the production and release of a diffusible inducer (PGF) which then activates homeotic genes, the products of which regulate those genes responsible for terminal differentiation.

1.5 Aims of Thesis

When *En-1* was first mapped close to or at the *Dh* locus it was hoped that *Dh* represented a mutant allele of *En-1* (Hill *et al* ,1987; Joyner *et al* , 1987) and it was this observation that prompted the work described in this thesis. The aims of this thesis were to determine if *Dh* and *En-1* were allelic and, if they were not, to devise ways to clone the *Dh* gene. Chapter three describes somes aspects of the phenotypic nature of the *Dh* mutation. Chapter four describes a linkage analysis performed to determine if these genes are allelic. Chapter five presents the results of a pulsed field gel electrophoresis analysis, a technique which allows the physical mapping of genes over a region of up to 10 megabases. The main aim of the work described in this chapter is to physically localise the *Dh* gene. Chapter six describes the construction of a gene targeting vector which will facilitate the actual cloning of the *Dh* gene.

Chapter Two

Materials & Methods

Unless stated otherwise all enzymes were supplied by Boehringer Mannheim Gmbh and all chemicals were supplied by BDH Ltd. All surgical procedures were performed under Home Office approval.

2.1 Anaesthesia

Mice were anaesthetised using one of three anaesthetics. Avertin or Hypnorm/Hypnovel was used for surgical manipulations such as tail tip removal and laparotomy whereas ether was used for tail bleeding.

2.1.1 *Avertin*

10mg tribromoethanol was added to 10ml tertiary amyl alcohol and then stored at 4°C until required.The actual anaesthetic was made up by dissolving 0.125ml of the stock solution in 10mls of pre-warmed (40°C) 0.9% saline.The dose was 0.02ml/gm mouse injected intra-peritonealy (i.p.).

2.1.2 *Hypnorm/Hypnovel*

Although Hypnorm (each ml contains 0.315mg fentanyl citrate and 10mg fluanisone; Janssen Pharmaceutica) and Hypnovel (midazolam hydrochloride; Roche) are both adequate anaesthetics on their own induction of anaesthesia is more reliable when they are used in combination. They cannot however be premixed and were therefore diluted 1:1 with water after which they were mixed (1 : 1) and stored at room temperature for two-four weeks.The dose was 0.2-0.3ml/30gm mouse injected i.p.

2.1.3 *Ether*

Mice were placed on a perforated "shelf" in a covered dessicator, the base of which contained cotton wool soaked in ether. Mice were carefully monitored in order to ensure that they were neither under- nor overexposed to the ether. Ether was routinely used to anaesthetise animals for tail bleeding, however even though tail bleeding is a relatively quick procedure ether is a very short acting anaesthetic and the mice were therefore continually exposed to small amounts of ether i.e. enough to maintain the anaesthesia. This was achieved by placing some cotton wool soaked in ether in the bottom of a 50ml Falcon tube and placing the open end of the tube at the mouse's head.

Mice were killed by cervical dislocation or carbon dioxide inhalation.

2.2 Surgical Procedures

2.2.1 *Tail Bleeding*

Mice were anaesthetised using ether. The tail veins were dilated by dipping the tail in warm water for approximately 1 minute (min).The end of the tail was placed in an eppendorf containing 15µl of 3.8% tri-sodium citrate (anti-clotting agent) and then nicked with a scalpel blade approximately 1/3 up the tail. The upper part of the tail was massaged in order to aid the flow of blood down the tail into the eppendorf. 100-200µl amounts were routinely collected and kept on ice until processed.The bleeding was stopped by firm application of cotton wool to the nicked tail. Mice were ear-clipped for future identification.

2.2.2 *Tail Tip Removal / Isolation of High Molecular Weight DNA*

Mice were anaesthetised and the terminal 1cm removed into ice cold 0.9% saline using a pre-heated scalpel blade. The tail tip was chopped up and placed in an eppendorf tube containing 700µl of homogenisation buffer (50mM Tris, pH 8.0, 100mM EDTA, 100mM NaCl, 1% sodium dodecyl sulphate). This was incubated overnight at 55°C in the presence of 35µl Proteinase K (10mg/ml in dH_2O) following which 20µl of RNase was added and this was then incubated at 37°C for a further 1-2 hours. The tube was then filled with phenol, shaken vigourously and then centrifuged for 15 minutes in a microfuge. The aqueous phase along with the interface was removed to a fresh tube and filled up with phenol:chloroform (1:1) shaken and centrifuged for 15 min. Again the aqueous phase along with the interphase was removed to a fresh tube and filled up with chloroform, shaken and centrifuged for 15 min. The aqueous phase was removed to a fresh eppendorf, this time leaving any interphase behind, filled up with isopropanol, and the tube was inverted several times until the DNA formed a stringy precipitate whereupon it was spooled out from the tube using a glass pasteur pipette (the pipette had been flamed to seal the end). Spooled DNA was then dipped in 70% ethanol, air dried and finally resuspended in 100-200µl of TE buffer (10mM Tris, 1mM EDTA). As the yield between samples varied the DNA concentration was checked by reading the optical density at 260 nm

2.3 Electrophoresis of Peptidase-3

Peptidase-3 is a red blood cell (rbc) enzyme, the different allozymes of which can be distinguished electrophoretically.

2.3.1 *Preparation of Haemolysate*

Haemolysates were prepared from washed, packed rbc. Blood from the tail was spun in a microfuge and the plasma removed. The rbc were then washed twice in 0.9% saline before adding 1/2 volume of distilled water. The lysate was then frozen and thawed twice to ensure complete lysis and adequate liberation of the enzyme.

2.3.2 *Electrophoresis*

Electrophoresis was performed on Titan III Cellulose 60x76mm acetate plates (Helena Laboratories, Gateshead, U.K.).These plates were pre-soaked in Tris-Borate-EDTA buffer (0.09M Tris, 0.05M Boric acid, 0.002M EDTA) for 20 min.The plates were removed from the TBE buffer immediately before use and gently blotted to remove any excess buffer. 10µl of haemolysate was loaded per well of the multiwell loading plate (Helena Labs) and then applied to the cathodal end of the cellulose plate using the Super Z applicator (1 application = 0.25µl).The plate was then placed sample side down across the zip-zone electrophoresis chamber (Helena Labs) and 200 volts were applied for 30 min.

2.3.3 *Detection of Peptidase-3*

Characterisation of peptidases is dependent on the detection of peptidase activity after electrophoresis.

Stain: 2ml 0.1M phosphate buffer pH 7.0 (2.28gm $K_2HPO_4.3H_2O$, pH to 7.0)

20µl MnCl 3.15gm/100ml

40µl *Crotalus adamanteus* snake venom 10mg/ml (Sigma)

40µl peroxidase 10mg/ml (Sigma)

80µl L-leucine-L-tyrosine 10mg/ml (Sigma)

80µl 4-amino-9-ethylcarbazole (1% in dimethyl formamide made up fresh just before use, Sigma)

Components added in the above order and then added to 2ml of 2% agar mixed gently and the entire mixture poured evenly over the electrophoresised cellulose plate and allowed to set.The plate was then wrapped in cling-film and incubated at 37°C for 1-2 hours. Peptidases stain up reddish brown on a yellow background.

2.4 Statistical Analysis

All calculations were performed on an Apple Macintosh Computer.

2.4.1 *Recombination Frequency (R.F.)*

R.F. was calculated as a percentage of the total number of offspring. The standard error (S.E.) was calculated as follows:

$$\text{R.F.} \pm \text{S.E.} = a/n \times 100 \pm \sqrt{\frac{a/n\ (1-a/n)}{n} \times 100}$$

Where a equals the number of recombinant animals and n equals the total number of animals.

2.4.2 *Chi Square Tests* (χ^2)

When comparing an observed proportion with a theoretical proportion; i.e. the number of Dh /+ offspring recovered compared to the expected recovery of 50%, the following standard formula was used:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

When comparing proportions, i.e. the number of Dh /+ offspring recovered from the DH cross compared to the DHF cross, the following standard calculation for a 2x2 contingency table with Yates correction was used:

$$\chi^2 = \frac{n\,(\,|ad-bc| - 0.5n\,)^2}{(a+b)\,(c+d)\,(a+c)\,(b+d)}$$

The vertical lines |ad-bc| means that the positive value of difference between ad and bc was used. The symbols are explained in Bailey (1959).

2.5 Bacterial Culture

2.5.1 *Media used for culturing E. coli.*

E coli. were normally grown on L agar plates or in L broth.

L broth: per litre:; 10g tryptone (Difco), 5g yeast extract (Difco), 10g NaCl, 2.46g $MgSO_4$.

L agar: per litre; L broth plus 15g agar.

H agar plates were used when the *lac* Z blue-white test was used to detect recombinant plasmids. This agar reduces the number of satellite colonies when plating out bacterial transformations. The substrate for the blue white test (BCIG/XGal) was also added to the agar.

H agar: per litre:; 10g tryptone (Difco), 8g NaCl. 0.02% BCIG (5-bromo-4-chloro-3-indoyl-b-D-galactoside .

"Terrific broth" (Tartoff & Hobbs,1987) yields high amounts of DNA and was used for large scale plasmid preparations.

Terrific broth: per 800mls; 12g tryptone, 24g yeast exytract, 4ml glycerol. 100mls of KPO_4 buffer (170mM KH_2PO_4 : 720mM K_2HPO_4) was added to the media before innoculation.

Bacteria were grown at 37°C with good aeration for liquid cultures. Antibiotics were used at the following concentrations: Ampicillin (Sigma) 100µg/ml, Kanamycin (Sigma) 30µg/ml. Stocks were prepared at 1000x concentration and stored at -20°C.

2.5.2 *Bacterial strains used*

All bacterial strains used were *E. coli* K-12. The most frequently used strain of *E. coli* used was JM83 (Messing,1983) as this strain is a host for pUC plasmids and consistently gave good transformation frequencies. The *lac* Z M15 gene is integrated into the host chromosome.

JM83: Δ *ara*, (*lac pro* AB), *rsp* L, F80, *lac* Z M15, (rk^+, mk^+)

Commercially bought MAX Efficiency DH5α competent cells (Gibco BRL) were also used (see Chapter 6).These cells have a high transformation efficiency for pUC based plasmids (Hanahan, 1983).

DH5α: F$^-$ø80d*lac* ZdM15, δ(*lac* ZYA-*arg* F)U169, *rec* AI, *end* AI, *hsd* R17(rk$^-$, mk$^+$), *sup* E44, λ- *thi* -I, *gyr* A, *rel* AI.

2.6 Vectors used for Cloning in Bacterial Cells

The pUC series of plasmids are based on pBR322 (Bolivar et al ,1977) have a high copy number in *E. coli* and carry the ampicillin resistance gene (Viera & Messing,1982). The pUC vectors have a multiple cloning site inserted into the β - galactosidase (*lac* Z) gene. Non-recombinant plasmids are able to synthesise this gene, which breaks down 5-bromo-4-chloro-3-indoyl-β-D-galactoside (Xgal), to release a blue indoyl derivative. In plasmids where an inset has been ligated into the polylinker the *lac* Z is interrupted and the colonies remain colourless.

2.6.1 *pTZ 18U/19U*

The pTZ plasmids are derived from the pUC plasmid (Mead et al ,1986) are 2.9 kb and contain the F1 origin of replication and therefore can be prepared as single stranded molecules. They also contain the universal and reverse primers for sequencing DNA.

2.6.2 *Bluescribe (Stratagene)*

This a 3kb plasmid derived from pUC 19 (Messing, 1983). As well as containing all of the standard features of pUC it has the T3 and T7 promoter sequences flanking the polylinker therefore facilitating the synthesis of specific single strand, sense and anti-sense, RNA transcripts.

2.7 Cosmid Library Screening

The library was screened using the method of Grunstein & Hogness (Sambrook et al, 1989). 200,000 colony forming units of the cosmid library were plated out onto a 20x20cm^2 Hybond membrane which had been placed upon H agar with kanamycin 30µg/ul. After incubation at 37°C for 12 hours two replica filters were taken from the master filter. The filters were asymmetrically marked with india ink, a process which enables the accurate orientation of autoradiographs to the master plate. All three filters were then placed on L agar plus antiobiotic and allowed to recover at 37°C for 3-4 hours. The replica filters were then processed by sequentially laying them on saturated 3MM Whatman filter paper in 20x20cm^2 dishes. They were first placed in 10% SDS for 2 min which causes membrane lysis. The filters were then denatured (0.5M NaOH, 1.5M NaCl) for 5 min, neutralised (1.5M Tris HCl, 1.5M NaCl, pH 5.5) for 5 min. and finally washed in 2xSSC to remove bacterial debris. The filters were then blotted dry and fixed by UV irradiation before being hybridised with the *En-1* specific cDNA λ4L7. The autoradiographs of the replica filters were compared. Duplicate positive signals were determined and the appropriate area (1cm^2) isolated from the master filter which had been stored at 4°C. The piece of filter was placed in 1ml of L broth / 15% glycerol and vortexed to release the colonies which could then be stored at -70°C indefinitely. Secondary screens were performed by plating out an aliquot (1x10^{-3}) of the primary pick onto gridded nitrocellulose filters in 9cm petri dishes. The

process of replicating, regrowing, denaturing and probing was carried out as before with the exception that nitrocellulose filters are fixed by baking for two hours at 80°C. Individual colonies could be isolated at this stage depending on the density of the colonies or if necessary a tertiary screen was performed.

2.8 Manipulation of Plasmid DNA

2.8.1 *Preparation of E. coli Competent for Transformation by DNA*

E. coli were made competent for DNA uptake by the method of Hanahan (1985). A fresh colony of JM83 cells was grown in L broth plus 10mM $MgCl_2$ to an O.D. of 0.5. The cells were then cooled on ice for 10 min. All subsequent steps were performed on ice. After pelleting at 3000 revolutions per minute (rpm) for 10 min the cells were resuspended in 0.3 volumes of RF1 (RF1 = 100mM rubidium chloride, 45mM manganese chloride, 30mM potassium acetate, 100mM calcium chloride, 15% glycerol, pH 5.8) and left on ice for 30 min. The cells were then pelleted and resuspended in 0.1 volumes of RF2 (RF2 = 10mM MOPS, 10mM rubidium chloride, 75mM calcium chloride, 15% glycerol, pH 6.8). After 15 min on ice 200µl aliquots were flash frozen in liquid nitrogen and stored at -70°C until required.

2.8.2 *Transformation of Competent JM83*

An aliquot of frozen competent cells was thawed on ice. Ligated DNA (<10µl) was mixed with the cells and left to adsorb for 20 min. The cells were then heat shocked at 42°C for 90 seconds (sec). 800µl of L broth plus antibiotic were added and the cells incubated for 45 min at 37°C. Aliquots of this transformation mix were then

plated out onto H agar containing ampicilin (100µg/ul) and 20µl Xgal (2% in dimethylformamide).

2.8.3 *Screening of Bacterial Colonies*

If a probe for the ligated insert was available the method of Grunstein & Hogness as described in section 3.6 was used to screen bacterial colonies. If no probe was available small scale plasmid preparations (mini-preps) were performed.

2.8.4 *Plasmid / Cosmid DNA Minipreparation*

A single plasmid containing colony of bacteria was added to L broth plus antibiotic and shaken at 37°C overnight. After pelleting at 3000 rpm for 10 min the cells were resuspended in 200µl of GTE (50mM glucose, 20mM Tris-Cl pH 8.0, 10mM EDTA) plus 10mg/ml lysozyme. After 10 min on ice 400µl of alkali / SDS (0.25M NaOH, 1% SDS) was added to denature chromosomal DNA. After a further 15 min on ice 300µl of 3M NaOAC was added to precipitate *E. coli* debris and chromosomal DNA. This was then centrifuged at 10,000rpm for 15 min. The supernatant was removed and the DNA precipitated with 0.6 volumes of isopropranol. The DNA pellet was washed in 70% ethanol, air-dried and resuspended in 30µl of TE. 3µl were used for digestion and eletrophoresis to estimate the concentration of DNA.

2.8.5 *Large Scale Preparations of Plasmid DNA*

250ml of Terrific broth plus appropriate antibiotic were innoculated with a single colony and grown overnight at 37°C. The culture was decanted into 250ml Sorvall centrifuge tubes and spun at 6000 rpm for 5 min. The cell pellets were resuspended in 20 ml of cold GTE (see above) and 10mg/ml lysozyme. After 15 min on ice 40ml of alkali / SDS (see above) was mixed in and incubated for a further 10 min after which 30ml of high salt (3M potassium acetate pH4.5 with gacial acetic acid) was added. This was left on ice for a further 15 min before being centrifuged at 12,000 rpm for 30 min. The supernatant was filtered through muslin and precipitated with 0.6 volumes of isopropanol. The DNA was pelleted by centrifugation (10,000 rpm, 20 min). The pellets were washed with 80% ethanol, lyophilised and redissolved in 3.2ml TE, 3.55g caesium chloride and 340µl ethidium bromide (10mg/ml, Sigma). The refractive index was adjusted, if necessary, to between 9.30 - 9.45. The sample was then centrifuged overnight at 40,000 rpm in a Sorvall Ultracentrifuge. Plasmid bands were removed with a syringe or a fine tipped pastette. The ethidium bromide was removed by a series of butan-2-ol extractions. DNA was then precipitated with 2.5 volumes of cold 75% ethanol. The DNA was spooled out, washed in 80% ethanol, dried and resuspended in TE.

2.9 Preparation of Mammalian DNA

2.9.1 *Preparation of High Molecular Weight DNA*

The liver of one adult mouse was homogenised in 10-15ml homogenisation buffer (1x buffer A1, 0.34M high grade sucrose, 2mM EDTA; Buffer A1 = 60mM KCl, 15mM NaCl, 0.15mM spermidine, 0.5mM spermine, 15mM Tris-HCl pH7.4).

After filtration through gauze the homogenate was carefully layered onto buffer A2, (1x buffer A1, 1.37M sucrose, 1mM EDTA), centrifuged in a swinging bucket rotor for 15 min at 12,000 rpm. The pellet was resuspended in 5ml buffer A1, mixed and then 5ml of 2x pronase buffer (2x = 100mM Tris pH10.0, 300mM NaCl, 200mM EDTA) was added. RNase A (100µg/ml) and SDS (final concentration of 0.2%) were added and the suspension was incubated at 37°C for 30 min following which the concentration of the SDS was increased to 1%. Proteinase K was added to a final concentration of 100µg/ml and the suspension was incubated overnight at 37°C following which it was phenol : chloroform extracted. The DNA was spooled out after the addition of cold ethanol and resuspended in TE.

2.9.2 *Preparation of DNA Agarose Blocks*

The procedure is the same as above, up to and including the centrifugation step. The pellet in this case was resuspended in buffer A3 (1x buffer A1, 1mM EDTA) at a concentration of 2-3x10^7 nuclei/ml. An equal volume of 1% molten low melting point agarose (LMP), dissolved in A3, was then added. The suspension was then mixed and 100µl aliquots were dispensed into the wells of a 96 well microtitre dish, the bottom of which had been removed and replaced with a removable plastic cover. The blocks were allowed to set at 4°C before being gently pushed out into 50ml NDS (0.5M EDTA, 1% lauryl sarcosine, 10mM Tris pH9.5) to which 10mg of proteinase K had been added. The blocks were then incubated at 50°C overnight following which the NDS / proteinase K was replaced and further incubated for 48 hours. The blocks were stored at 4°C in fresh NDS until required.

2.9.3 *Preparation of Yeast Chromosome Markers*

A single colony of yeast cells was grown to the late log phase (approximately 10^8/ml) in 100ml YPD (YPD per 500ml : 5g bacto yeast extract (Difco), 10g bacto peptone (Difco) made up to 450ml with H_2O. 50ml of 20% glucose added after sterilising both solutions). After harvesting the cells were washed twice in 0.05M EDTA pH 7.5 and centrifuged. The cell pellet was resuspended in 3ml 0.05M EDTA and mixed with 5ml of LMP agarose (prepared in 0.125M EDTA pH 7.5) at 42°C. 1ml of SCE (1M sorbitol, 0.1M sodium citrate, 0.06M EDTA pH 7.0), 0.5ml β-mercaptoethanol and 10mg zymolyase 60,000 (Miles) were added to the suspension. 100µl aliquots were then dispensed into a 96 well microtitre dish (see above). The agarose blocks were incubated overnight at 37°C in 0.45M EDTA pH 9.0, 10mM Tris pH 8.0, 7.5% v/v β-mercaptoethanol in order to generate spheroplasts. The blocks were then incubated overnight at 50°C in NDS and 10mg proteinase K. The blocks were stored at 4°C in fresh NDS until required.

2.9.4 *Preparation of λ concatemers*

25ml of L broth plus 0.01M $MgSO_4$ was innoculated with a single colony of λ containing bacteria and incubated overnight at 33°C. 6ml of this overnight culture was then added to each of two 250ml volumes of L broth plus 0.01M $MgSO_4$, shaken at 33°C for 3 hours (until $OD_{600} = 0.45$), induced at 43°C for 15 min and then shaken at 39°C for 2.5 hours. The culture was then centrifuged at 8,000rpm for 10 min and the pellet resuspended in 8.8ml TE / 0.5ml $CHCl_3$ and shaken at 37°C for 15 min. 20µl of RNase A and DNase I (both at 10mg/ml) were added and the incubation continued for a further 15 min followed by centrifugation at 10,000rpm for 15 min. The supernatant (9.2ml) was removed, mixed with 9ml 1% LMP agarose and aliquoted into a 96 well

microtitre dish (see above). The blocks were incubated in NDS / proteinase K (10mg) at 50°C overnight and then stored at 4°C in fresh NDS until required.

2.10 Manipulation of DNA by Enzymes

2.10.1 *Restriction Endonuclease Digestion*

Restriction endonucleases were purchased from Boehringer Mannheim Gmbh., Bethesda Research Laboratories and New England Biolabs. Digests were performed according to the manufacturers specifications. Reactions were stopped by heat inactivation, phenol extraction, freezing at -20°C or by the addition of stop mix (15% ficoll, 0.25M EDTA and orange G to colour).

2.10.2 *Restriction Endonuclease Digestion of DNA Blocks*

The blocks were prepared for digestion by soaking in 5ml TE and 5µl PMSF (20µg/ml in isopropanol) at 50°C for 30 min. The blocks were then washed twice in TE at 4°C for 30 min following which they were equilibrated twice in 1ml restriction buffer at 4°C for 30 min. The block was then digested with 20-30 units of enzyme in 100-200µl restriction buffer overnight at the appropriate temperature.

2.10.3 *Ligation of DNA Molecules*

Bacteriophage T4 DNA ligase was used for all ligation reactions as this ligase can ligate both cohesive and blunt ended DNA fragments. Ligations were performed at 16°C overnight with 1 unit of ligase in 50mM Tris-HCl pH 7.4, 10mM $MgCl_2$, 1mM spermidine, 100µg/ml bovine serum albumin (BSA), 1mM adenosine triphosphate (ATP) and 1mM dithiotreitol (DTT). The ratio of vector to insert was varied according to the insert size and whether the reaction was ligating cohesive or blunt ended termini.

2.10.4 *Filling In Cohesive DNA Termini*

Bacteriophage T4 DNA polymerase has a 5' -> 3' polymerase activity and a 3' -> 5' exonuclease activity, which is more active on single strand DNA than on double strand DNA, and is therefore particularly useful for 'blunt-ending' cohesive DNA termini. Reactions were carried out at 37°C for 30 min with 1-2 units of T4 DNA polymerase in 1x T4 DNA polymerase buffer (10x = 330mM Tris acetate pH 8.0, 660mM potassium acetate, 100mM magnesium acetate, 5mM DTT, 1mg/ml BSA) and 2mM dNTP. The high concentration of dNTP directs the reaction towards the polymerization (filling in) reaction.

2.10.5 *Enzymatic Phosphorylation of Non-phosphorylated Linkers*

Only phosphorylated molecules are substrates for bactriophage T4 DNA ligase. Non-phosphorylated linkers must be treated with bacteriophage T4 DNA kinase and ATP before they can be ligated to DNA. 0.5-2.0µg of non-phosphorylated linkers were mixed with 1µl 10mM ATP, 1µl 10x kinase buffer (10x = 0.66M Tris-HCl pH7.5, 0.1M $MgCl_2$, 100mM DTT) and distilled water in a total volume of 10µl. Two

units of bacteriophage T4 polynucleotide kinase were then added and incubated for one hour at 37°C after which the linkers could be used without further purification.

2.10.6 *Ligation of Phosphorylated Linkers to Blunt Ended DNA*

0.1 - 0.5µg of blunt end DNA (in a volume of 7µl or less), 1-2µg of phosphorylated linkers (in a volume of 8µl or less), dH_2O to 15µl, 2ul of 15mM ATP, 2µl 10x blunt end ligation buffer (0.66M Tris-HCl pH 7.6, 50mM $MgCl_2$, 50mM DTT, 5mM spermidine) and 1µl of bacteriophage T4 DNA ligase were mixed together in the above order and incubated overnight at 16°C. Unligated linkers could be removed by passing the ligation mix through a SizeSelect-400 Spun Column (Pharmacia)

2.11 Gel Electrophoresis

Electrophoresis through agarose is a standard method which allows the separation, identification and purification of DNA fragments. DNA molecules from 200bp to approximately 50kb can be resolved on standard agarose gels of various concentrations, run in an electric field of constant strength and direction. DNA fragments up to 10 megabases in length can be resolved by pulsed field gel electrophoresis (PFGE) in which the direction of the electric field is periodically alternated.

2.11.1 *Conventional Gel Electrophoresis (resolution of DNA fragments up to 50kb)*

Horizontal agarose gels were poured in a variety of sizes depending on the number of samples being examined. Agarose (Sigma, Type II) concentrations ranged from 0.4% to 1.5% depending on the size range to be resolved. The gels were made up and run in E buffer (0.04M Tris, 0.02M sodium acetate, 0.001M EDTA, pH 8.2). DNA samples were mixed with a 1/10 volume of stop mix (15% ficoll, 0.25M EDTA and orange G to colour) which acts as a visible marker for the migration front of small (200-500bp) DNA fragments. Gels were run at voltage gradients of up to 10V/cm then stained with a few drops of ethidium bromide (10mg/ml) for 10-15 min. Size markers were ØX174, digested with *Hae* III and lambda, digested with *Hind* III (lambda digested with *Sma* I and *Eco* RI was also used for sizing cosmids). Gels were then photographed with overhead illumination from short wave (254nm) UV lamps (Mineralight). Kodak Technical Pan film was used with a Polaroid MP4 land camera fitted with a red filter (Hoya 25A). Films were developed in a Fuji RGB-2 Automatic Film Processor.

2.11.3 *Pulsed Field Gel Electrophoresis*

Pulsed field gel electrophoresis (PFGE) was used to resolve DNA molcules greater than 50kb. The PFGE apparatus (Waltzer) of Southern *et al* (1987) was used in all experiments. One percent agarose gels were made up in 0.5x TBE (1x = 1M Tris, 0.8M boric acid, 20mM EDTA, pH8.3). The gels were run at 5V/cm for 24-48 hours. The pulse times were adjusted depending on the size of fragments to be resolved. The buffer was circulated through a cooling system in order to mainain a constant temperature of 10-12°C. Size standards used were oligomers of lambda

bacteriophage DNA and chromosomes of *S. cerevisiae*. The gels were photographed as described above.

2.11.3 *Preparative Agarose Gel Electrophoresis*

Specific DNA fragments for cloning or radiolabelling were isolated from low melting point (LMP) agarose gels. LMP agarose gels at 30°C and melts at 65°C, a temperature which is well below the melting temperature of most double stranded DNA. After electrophoresis the gel was visualised with long wave length (300nm) UV light which minimises damage to the DNA and the gel slice containing the DNA was excised. The DNA was recovered via Spin-X columns (Costar) or Geneclean II (BIO 101 Inc). Both of these methods routinely yielded high quality DNA which could be used without further purification.

2.12 Southern Blotting

Southern blotting is a technique which allows the transfer of DNA fragments from agarose gels to a solid support (nitrocellulose filters or nylon membranes).

After making a photographic record the gel was denatured in 500mM NaOH, 1.5M NaCl for 45 min followed by neutralisation in 1M Tris, 2M NaCl for 45 min. Pulsed field gels were subjected to depurination in 0.25M HCl for 20 min before being denatured. Depurination breaks down large DNA fragments into sizes of approximately 1kb which can then be transferred rapidly from the gel with high efficiency. Denatured and neutralised gels were placed on a support over Whatman 3MM paper soaked in 20x SSC (3M NaCl, 0.3M sodium citrate, pH 7.4). The sides of the gel were sealed with cling film in order to prevent side diffusion and / or

evaporation. Nitrocellulose (Schleicher & Schuell) or Hybond-N (Amersham) cut to size and pre-soaked in distilled H_2O was carefully laid over the gel. Air bubbles were removed and two pieces of Whatman filter paper cut to size were laid on top followed by several layers of paper towels. A weight was then placed on top to facilitate capillary transfer of the DNA to the membrane. Hybond was routinely used for the transfer of DNA as this is much more durable than Nitrocellulose. The DNA was normally transferred overnight. Pulsed field gels were transferred for 48 hours with the replacement of fresh paper towels after 24 hours. The lanes of the gel were marked on the membrane which was then baked at 80°C for 2 hours (Nitrocellulose) or UV irradiated for 3 min (Hybond). The filters could then be kept at room temperature until required for hybridisation.

2.13 Preparation of Radiolabelled DNA Probes for Hybridisation

2.13.1 *Nick Translation*

E. coli DNA polymerase I (Pol I) can incorporate nucleotides to the 3' hydroxy terminus of DNA which has been nicked by DNAse I. Pol I also has a 5'-> 3' exonuclease activity which can remove nucleotides from the 5' side of the nick. By replacing pre-existing nucleotides with radioactive nucleotides, (^{32}P α-CTP), radiolabelled DNA can be made via the simultaneous removal of nucleotides from the 5' side and the addition of nucleotides to the 3' side in the movement of the nick along the DNA.

A BRL Nick Translation Kit was used according to the manufacturers recommendations. Incorporated nucleotides (labelled DNA) were isolated from unincorporated nucleotides using Pharmacia Nick Translation Columns. Incorporation of radioactivity was assessed via TCA (trichloroacetic acid) precipitation which

selectively precipitates molecules > 20 nucleotides. Probes were denatured with 0.2M NaOH before being added to the hybridisation bag.

2.13.2 *Random Prime DNA Labelling*

This method of generating radioactive DNA probes is particularly useful when small amounts of DNA are available (e.g. 10ng). Denatured DNA serves as a template for random hexanucleotide primers. The 3' hydroxy terminus of the hexanucleotide then serves as a primer for the synthesis of DNA via the Klenow fragment of DNA polymerase I. Addition of radioactive nucleotides (^{32}P α-CTP) generates a radiolabelled probe.

A Boehringer Random Prime kit was used according to the manufacturers recommendations. Incorporated nucleotides were isolated from unincorporated nucleotides using Pharmacia Nick Translation Columns. Radioactivity was assessed via TCA precipitation. Probes were denatured with 0.2M NaOH before being added to the hybridisation bag.

2.13.3 *End Labelling of Oligonucleotides*

Synthetic nucleotides are synthesised without a phosphate group at the 5' terminus. T4 polynucleotide kinase catalyses the transfer of γ^{32}P from (γ^{32}P)ATP resulting in radioactive end labelled probes.

30picomoles of DNA oligonucleotides were incubated in 1x T4 polynucleotide kinase buffer, (10x = 0.5M Tris pH7.6, 0.1M $MgCl_2$, 50mM DTT, 1mM spermidine, 1mM EDTA pH 8.0), 25µCi (γ^{32}P)ATP and 1 µl of polynucleotide kinase in a volume of 20µl for 30 min at 37°C. Incorporation was assessed via chromotography on DEAE

(Whatman) paper in 3M ammonium formate - unincorporated nucleotides move through the paper with the solvent front whereas incorporated nucleotides remain at the point of application. The probes were ready to be used without further treatment.

Oligonucleotides were synthesised and purified according to the manufacturers instructions on the Applied Biosystens 381A Oligonucleotide Synthesiser.

2.14 Hybridisation

Filters were pre-hybridised and hybridised in Quick-Hyb (5xSSC, 0.1% BSA, 0.1% ficoll, 0.1% polyvinylpyrrolidine, 0.5% SDS, 0.5% sodium pyrophosphate and 100µg/ml denatured salmon sperm DNA). Filters were sealed in plastic bags and pre-hybridised at 68°C for at least 30 min after which denatured probe was carefully added to the bag which was then resealed and hybridised at 68°C overnight. The bags were opened in wash solution (2xSSC, 0.1% SDS, 0.1% sodium pyrophosphate) at the hybridisation temperature and the filters removed. This wash solution was removed, replaced with fresh wash, and incubated at 68°C for 20 min. Filters were washed at the appropriate stringency depending on the nature of the probe. Filters were then wrapped in clingfilm and autoradiographed overnight at -70°C with pre-flashed X-AR Kodak film. Films were developed in a Fuji RGB-2 Automatic Film Processor.

The hybridisation conditions were altered when oligonucleotide probes were being used. The hybridisation temperature was calculated according to the base composition of the oligonucleotide:

$$4\,(G+C) + 2\,(A+T) + 10 = \text{Temp } ^{o}C$$

Post hybridisation washes were carried out at the hybridisation temperature in 4xSSC, 0.1%SDS, 0.1% sodium pyrophosphate.

Chapter Three

The Dh Phenotype

3.1 Introduction

The study of the developmental effects of mutant genes in the mouse is important for two main reasons. Firstly it will aid the our understanding of how genes control development by investigating how the mutant gene disrupts normal development. Secondly as the mouse is a representative vertebrate and mammal and since its genetic constitution is better known than any other mammal it will help give an insight into basic mammalian development and physiology.

Mammalian embryos are not easily accessible to manipulation and it is extremely difficult to interfere with the normal development of the mouse embryo although the recent advances in homologous recombination and ES stem cell chimeras have made this more possible (reviewed by Capecchi, 1989). However the vast numbers of mutant lines already available in the mouse give the researcher a huge pool whereby the mutated gene and the resultant consequences are available for study. Mice which carry the *Dh* mutation fall into this category.

As described in the general introduction, the *Dh* mutation has a pleiotropic phenotype which includes pre-axial hind limb abnormalities, asplenia, a decreased number of pre-sacral vertebrae and ribs, a small and defective digestive system and urogenital abnormalities. The major purpose of generating *Dh* /+ mice in this study was to determine whether *Dh* and *En-1* were allelic rather than to study the mutation at an embryological / anatomical level. However it is possible to comment on a few aspects of the phenotype.

3.2 Hind Limb Abnormalities

Classification of all offspring generated in this study, with regard to whether they carrried the Dh gene, was achieved initially by the examination of the hind limbs. Searle (1964) had previously reported that the expressivity of the gene with regard to the hind limb abnormalities was variable and included:

1) Slight thickening / lengthening of the hallux (big toe),

2) Polydactlyly pre-axial to the hallux (i.e. extra toe anterior to the limb axis),

3) Oligodactyly (loss of digits) and

4) Luxation (dislocation) and reduction in the length of one or both hind limbs.

The limb abnormalities exhibited by the mice generated in this study certainly confirm the variable expression and included all of the above (see Figs. 3.1, 3.2), although the most common abnormality was a lengthening of the hallux. Although the long bones of the hind limbs were not directly examined, Fig. 3.3 shows three mice all of which had long bone abnormalities; this again showed a variable expression. The abnormalities of either the digits or the longbones were not necessarily the same on both limbs. Figure. 3.4 are drawings of some of the hind limb abnormalities found in *Dh* /+ and *Dh/Dh* animals as described by Searle.

In the general introduction it was proposed that the anterior (cranial) shift in the hind limb bud impaires a critical early developmental relationship between the limb bud and somites. Rooze (1977) proposed a double action model to explain the limb abnormalities of *Dh* mice. The first event to occur is the anterior displacement of the hind limb bud. He then proposed that in +/+ mice the adjacent somites have a graded morphogenetic influence on the limd bud, whereby the influence of the posterior (caudal) somites is greater than that of the anterior (cranial) somites. In *Dh* mice he proposed that the graded morphogenetic influence of the somites extends anteriorly.

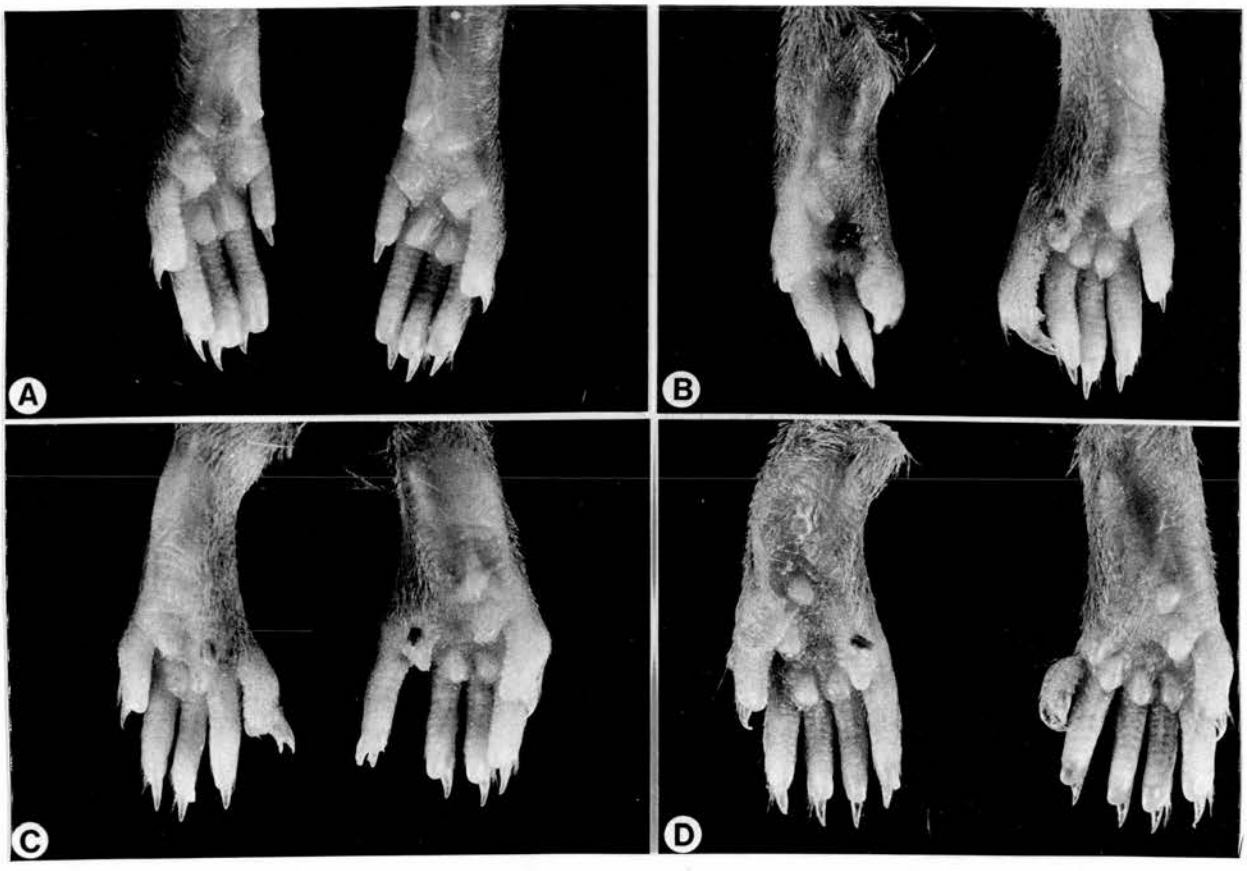


Figure 3.1. Underside view of hind feet (left and right refer to left and right of mouse). *A*, Normal (+/+) hind feet, *B*, *C*, and *D* are from *Dh* /+ mice. *B*, Oligodactyly of left foot, lengthening of right hallux. *C*, Thickening / lengthening of left hallux, lengthening of right hallux, note the abnormal nail on both affected digits. *D*, Lengthening of left hallux, polydactyly preaxial to the right hallux.

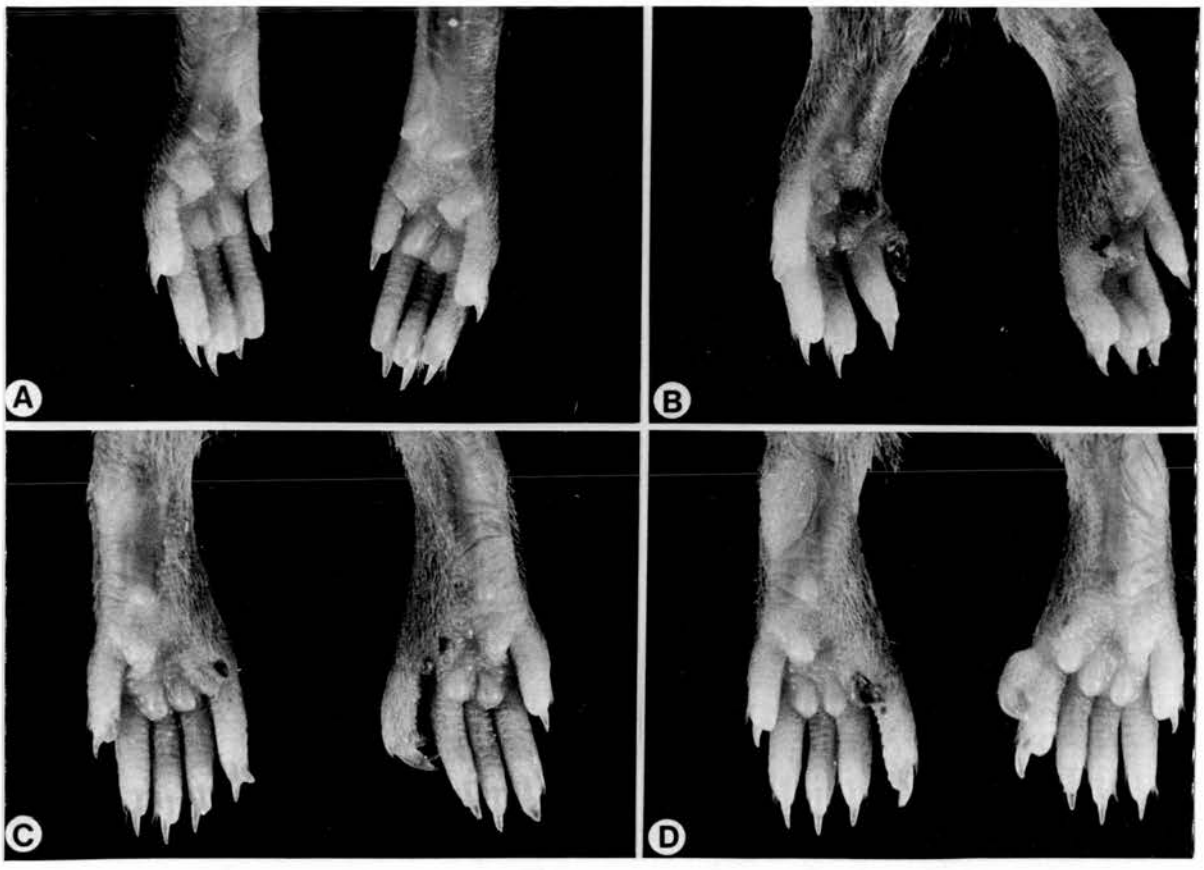


Figure 3.2. Underside view of hind feet (left and right refer to left and right of mouse). *A*, Normal (+/+) hind feet, *B*,*C* and *D* are from *Dh* /+ mice. *B*, Both feet exhibit loss of hallux (oligodactyly). *C*, Both feet exhibit lengthening of hallux. The left foot also has an abnormal nail. *D*, Lengthened left hallux, the right foot has polydactyly preaxial to a lengthened hallux.

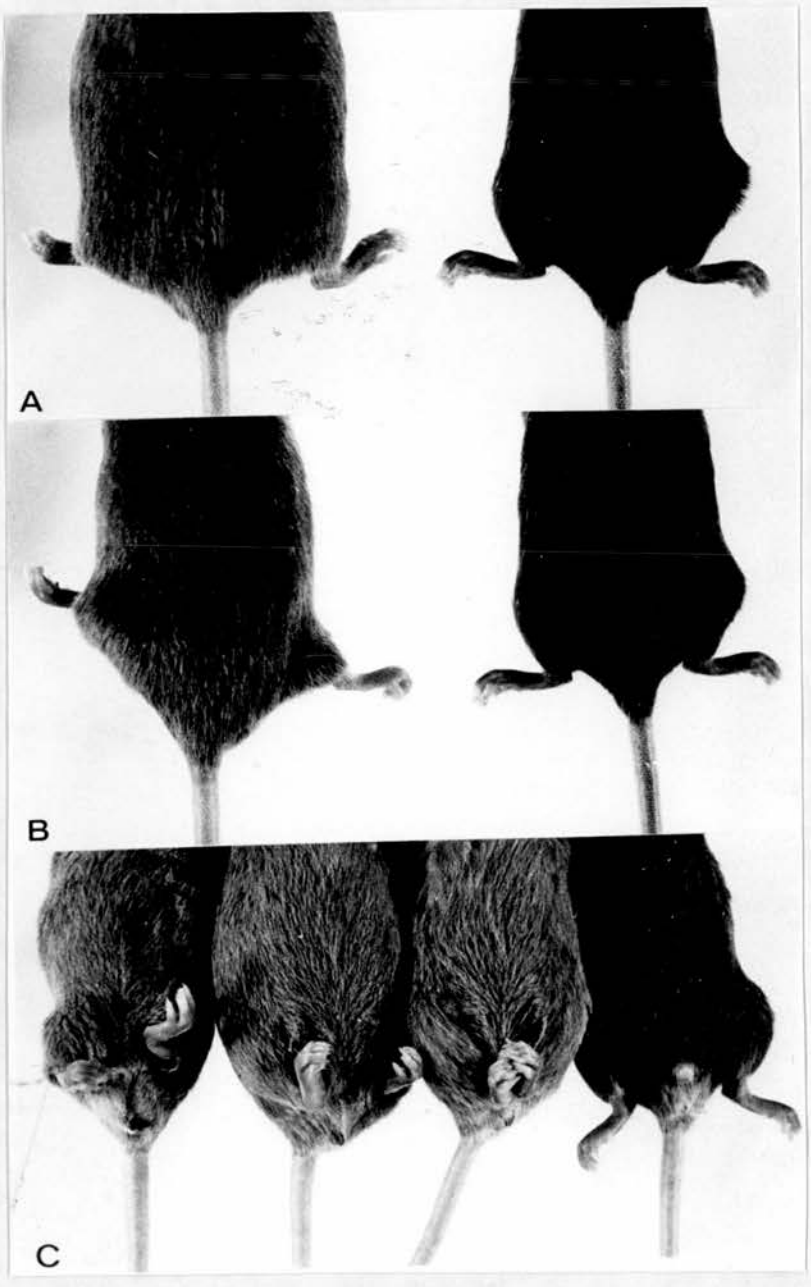
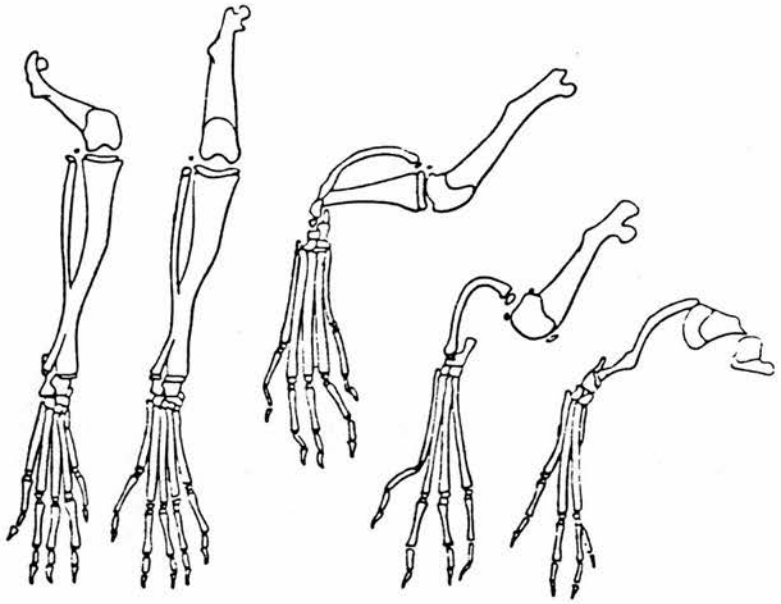
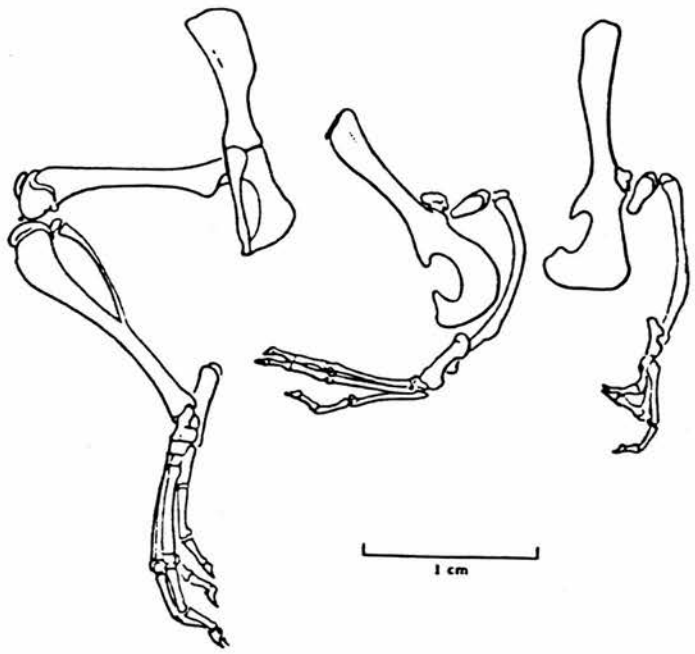


Figure 3.3. *A* and *B* show dorsal view of *Dh* /+ (left) and +/+ (right) mice. *C*, is a ventral view of three *Dh* /+ (left) and one +/+ (far right) mouse. Note the variable expression of the abnormality with regard to the long bones and that the abnormalities are not necessarily the same on both limbs.



A



B

Figure 3.4. *A*, Right hand limb skeletons of adult normal (left) and *Dh* /+ mice showing different grades of abnormality in the latter, from preaxial triphalangy to severe oligodactyly, with loss of tibia and fragmentation of the femur. *B*, Right hand limb skeltons of normal (left) and surviving *Dh* / *Dh* mice, with severe preaxial defects in the latter. (Reproduced, with permission, from Searle, 1964).

These two genetic effects would lead to different abnormalities according to whether the anterior shift of the limb bud is negligible, moderate or important. For example, if the limb bud displacement is negligible, the amount of pre-axial mesoderm remains the same but the anterior shift of the strong morphogenetic influence of the somites causes excessive growth of the hallux. In a massive displacement of the limb bud the morphogenetic influence of the somites cannot compensate and more severe abnormalities occur. However the major problem with this 'double action' model is that, when proposed, the anterior extension of the stronger morphogenetic influence of the caudal limb adjacent somites was completely hypothetical. To date there is no evidence to support the idea about a graded morphogenetic influence of somites on limb development.

Although exactly how the limb is patterned is not clear, it is interesting to note that it has recently been shown that *En-1* protein is present in the ventral ectoderm of the developing mouse and chick limb buds (Davis *et al.*, 1991). This conservation of expression strongly suggests that *En-1* has a function in the developing limb bud and may be involved in pattern formation. It would be of great interest to determine the expression pattern of *En-1* in *Dh* mice.

3.3 Asplenia

Asplenia is a constant feature of the *Dh* mutation in both the hetero- and homozygous animals. On the basis of spleen classification Searle found the *Dh* mutation to be 100% penetrant but only 96% penetrant on the basis of hind limb abnormalities. As the initial classification of *Dh* /+ mice in this study was based on hind limb abnormality the above observations would suggest that some *Dh* /+ animals would go undetected. However, apart from those recombinant animals kept alive in

order to maintain the recombinant chromosome, all offspring were checked for the presence or absence of spleen (approximately 550 animals). There was found to be complete concordance with regard to hind limb abnormalities and asplenia.

3.4 Kidney Abnormalities

In his study of the *Dh* mutation Searle found that 6/25 (24%) *Dh* /+ animals had a hydropic left kidney whereas 19/25 (76%) had an abnormal flattening of the antero-ventral region of the left kidney where it would normally be in close contact with the spleen. In homozygous animals the uro-genital abnormalities were much more severe and included recto-vaginal fusion in 40% of females, recto-urethral fusion 70% of males, absent / vestigial bladder 32% and blind ending ureters (20%), urethras (20%) or uterine horns (10%).

Of the offspring generated in this study 5/563 (0.9%) had a hydropic left kidney (see Fig. 3.5). This condition presumably results from a physical obstruction or a kink in the left ureter. Ureteric kinks have been described in mice carrying the luxate gene (Carter,1953). Luxate (*lx*) mice have a great similarity to *Dh* mice particularly with respect to the loss of pre-sacral vertebrae, shortening of the coelom and range of hind-limb abnormalities. Ureteric kinks implies that the ureter is longer than the region in which it has to fit. The urogenital system develops in the mesoderm of the posterior coelom a region known to be defective in *Dh* /+ animals. In the homozygous form it may be that although the kidney develops normally its spatial orientation is impaired. However it must be pointed out that the kidney abnormalities described here were only detected when the mice were sacrificed. The incidence of hydropic left kidney was significantly lower than Searle's findings ($\chi^2 = 57.63$, $P<0.001$). This discrepancy may reflect differences in genetic background or age at autopsy. Most of the *Dh* /+ mice in this study were sacrificed at weaning age.

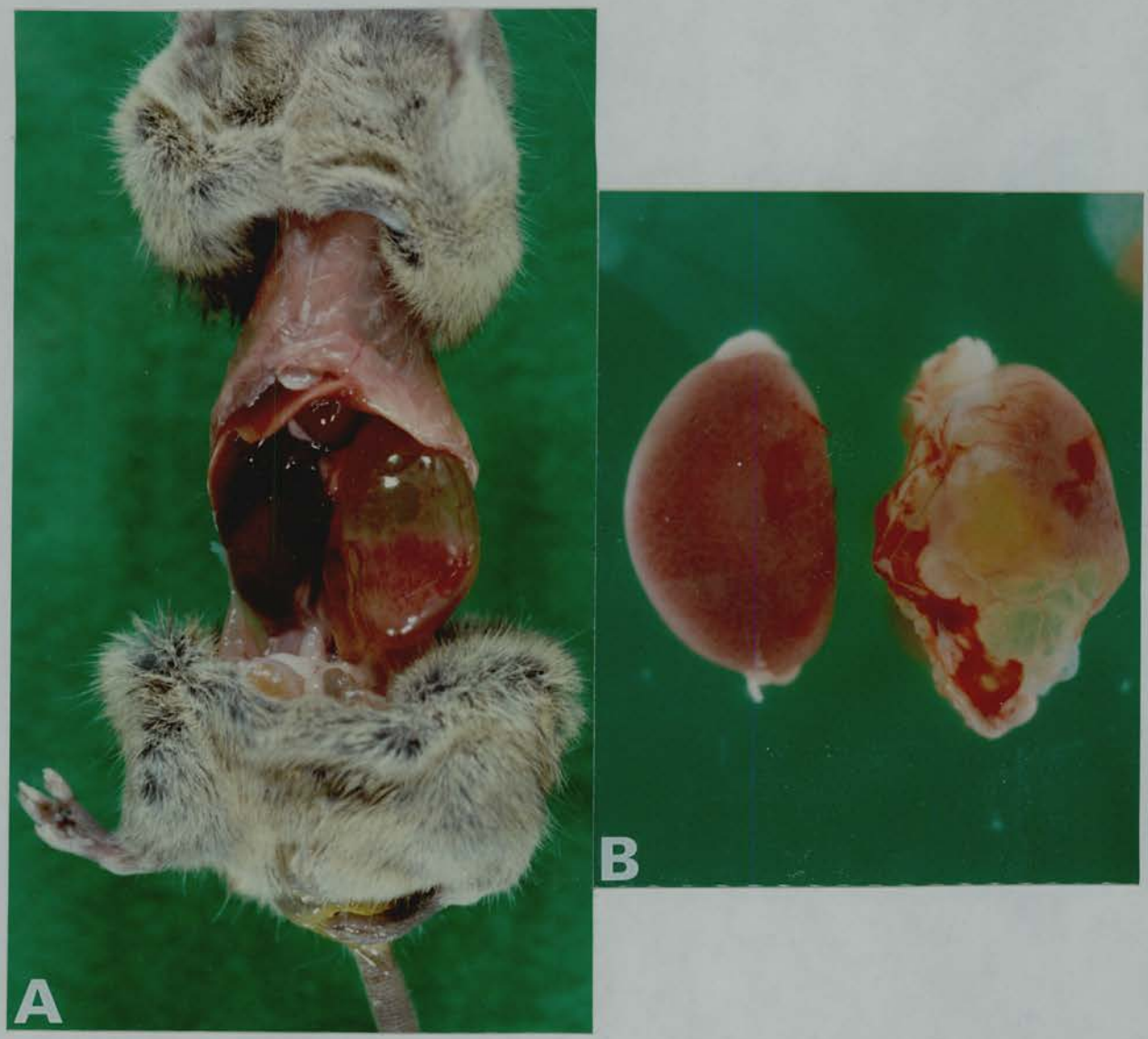


Figure 3.5. *A*, Hydropic left kidney of *Dh* /+ mouse (intestine removed). *B*, Comparison of the hydropic left kidney with the normal right kidney. The morphology of a large portion of the left kidney has been severely affected. Note that there has been extensive fluid loss from the hydropic kidney upon removal.

3.5 Recovery of the Dh Mutation

Heterozygous *Dh* /+ mice were outcrossed to wild type (+/+) mice, 50% of the offspring were expected to be heterozygous for *Dh*. The data are summarised in Table 1. *Dh* /+ mice were classified via limb morphology and asplenia at autopsy.

Table 1. Inheritance of the Dh Mutation and Expectations for Autosomal Inheritance Assuming all Heterozygotes are Viable.

DH Cross**	Progeny			Observed	Expected	Significance [a]
Parent	Dh/+	+/+	Total	%Dh/+	%Dh/+	χ^2
Dh/+ Female	13	39	52	25.0	50	13.00*
Dh/+ Male	65	165	230	28.3	50	43.48*
Total	78	204	282	27.7	50	56.30*

DHF Cross**	Progeny			Observed	Expected	Significance [a]
Parent	Dh/+	+/+	Total	% Dh/+	%Dh/+	χ^2
Dh/+ Female	69	113	182	37.9	50	10.64*
Dh/+ Male	33	66	99	33.3	50	11.00*
Total	102	179	281	36.3	50	21.10*

Combined	Progeny			Observed	Expected	Significance [a]
Parent	Dh/+	+/+	Total	%Dh/+	%Dh/+	χ^2
Dh/+ Female	82	152	234	35.0	50	20.94*
Dh/+ Male	98	231	329	29.8	50	53.77*
Total	180	383	563	32.0	50	73.20*

a. χ^2 tests significance of departure from Mendelian expectation if all heterozygotes are viable.

* $P < 0.001$

** DH and DHF crosses explained in chapter four.

From the data presented in Table 3.1 it can be seen that without exception the number of *Dh* /+ animals recovered from either cross is significantly different from the numbers one would expect assuming that all *Dh* /+ animals are viable. This appears to contrast greatly with the study performed by Searle (1964) who found that although fewer *Dh* /+ mice were recovered (44.7%) than expected in a standard outcross analysis the difference was not significantly different from the expected 50%. The animals in this study were scored at weaning whereas Searle scored for *Dh* at birth. Searle also reported that the survival rate to weaning for *Dh* /+ animals was 73%. If one assumes 44.7% recovery at birth followed by 73% survival to weaning then one would expect approximately 32.6% of all *Dh* /+ animals born to survive to weaning. Similarly if the frequency of *Dh* was 50% at birth the expected weaning frequency would be 36.5%. The observed frequency at weaning was 180/563 (32.0%) overall which is not significantly different from 32.6% (χ^2 = 0.10, P>0.05) but is significantly different from 36.5% (χ^2 = 4.98, P<0.05). The recovery rates for the *Dh* /+ animals in the DH and DHF cross are 27% and 36% respectively neither of which are significantly different from 32.6%. Thus it would seem that although the recovery of *Dh* /+ animals at weaning is slightly lower than one would expect this difference may be accounted for by a small deficiency at birth plus a reduced viability to weaning.

Chi square tests were performed in order to determine whether any significant differences in recovery of *Dh* /+ animals occured either between or within each cross. When comparing between crosses, i.e. either DH female to DHF female (χ^2 = 2.42, P>0.05) or DH male to DHF male (χ^2 = 0.63, P>0.1), no significant differences were found. Likewise no significant differences were found when comparing between sexes within each cross: DH female : DH male (χ^2 = 0.09, P>0.05) ; DHF female : DHF male (χ^2 = 0.40, P>0.05) When comparing the two crosses en masse there is a slight difference (χ^2 = 4.44 , P< 0.05).

With regard to parental origin, the number of *Dh* /+ animals recovered in either cross shows no significant difference therefore suggesting that parental transmission is of no consequence with regard to the recovery of *Dh*. As noted above, the expressivity (severity of the limb abnormality among *Dh* /+ mice), was variable. Although no systematic records were kept on the degree of severity, individual litters from *Dh* /+ x +/+ crosses often included both severe and mild forms of limb abnormalities. This suggests that it is unlikely that the expressivity of the *Dh* gene is rigidly related to parental transmission and therefore unlikely to involve genomic imprinting.

3.6 Discussion

In the general introduction it was proposed that the primary effect of the *Dh* gene was affecting segmentation, i.e. somitogenesis, and that the normal organised structure of the splanchnic mesoderm is defective as a result. Although it is possible to construct a feasible chain of events to explain the pleiotropic effect of *Dh* the pleiotropism could, instead, be the result of a chromosomal deletion which, as well as deleting the *Dh* gene, is also deleting independent but closely linked genes. Alternatively it may also be that the *Dh* gene is affecting each tissue independently. The existence of genetically independent loci, i.e. luxate (*lx*) and luxoid (*lu*), which exhibit similar pleiotropic phenotypes, would suggest that the pleiotropism seen in *Dh* /+ animals is the result of a single gene effect rather than the deletion of several linked genes. *Dh* and *lx* have been shown to interact with each other genetically (Searle, 1964). Animals which were heterozygous for both genes, show a high incidence of oligodactyly and tibial defects. Homozygosity for *lx* enhances the visceral abnormalities of *Dh*. Likewise *lu* and *lx* interact with each other with regard to the limbs defects with a cumulative effect in double heterozygotes (Forsthoefel, 1958, 1959; Green, 1955). The similar phenotypes of these genes, i.e. disruption in number of pre-sacral vertebrae, pre-axial limb defects, urogenital abnormalities, abnormal coelom, and their interaction with each other would imply that they are affecting the same developmental pathway and therefore unlikely to be independently affecting different tissues. Although the spleen anomaly is unique to *Dh,* oligodactyl mice (*ol*) which have abnormalities of all four limbs (postaxially, i.e. on the posterior side of the limb axis), the skeleton and the urogenital system also have a small and sometimes absent spleen (Freye, 1954).

It is worth noting that, in general, preaxial abnormalities tend to be confined to or more severe in hind limbs (*Dh* Searle, 1964; *lx* Carter, 1951; *lu* Forsthoefel,

1958). Postaxial abnormalities predominantly occur in fore limbs (*px* Searle, 1964; *ol* Freye, 1954). There is no apparent relationship between the loss or gain of vertebrae and the side of limb affected, e.g. *Dh* and *lu* mice both exihibit preaxial abnormalities of the hind limb but *Dh* mice have fewer presacral vertebrae whereas *lu* mice have more. The reasons for the above observations remain unclear.

Asplenia is a constant feature of *Dh* and this aspect of the mutation has been extensively exploited for various immunolgical studies. The mouse spleen, which develops from a localised thickening of the anterior splanchnic mesodermal plate (ASMP see page 21; Theiler, 1972; Green, 1967), lies in the upper left part of the abdomen and is composed of two types of tissue. The red pulp is mainly concerned with the destruction and removal of effete erythrocytes whereas the white pulp houses the T and B cells of the immune system. The spleen is a major peripheral lymphoid organ which together with the lymph nodes traps antigen, houses proliferating lymphocytes following their stimulation by antigen, and subsequently releases lymphoid cells and their products, e.g. antibody, into the body. Thus the spleen, although not vital to sustain life, has a major role to play in the immune system and the effect of asplenia on this system has been well studied. In the majority of these studies animals carrying the *Dh* mutation are the animal model of choice.

In *Dh* /+ animals it has been shown that, although histologically normal, the lymph nodes are enlarged with twice as many cells present compared to +/+ siblings (Lozzio, 1972; Lozzio & Machado, 1975). This is probably a compensatory effect of asplenia as the murine spleen normally houses a large pool of lymphocytes. The serum levels of IgM and IgG_2 are decreased in *Dh* /+ animals (Battisto *et al.*, 1969; Lozzio & Wargon,1974) although cellular immunity, as based on spleen and skin allograft rejection times, was found to be normal (Searle, 1964; Lozzio & Wargon, 1974). However this does not mean that some aspects of T cell function are not impaired.

Reconstitution of irradiated +/+ siblings with various combinations of B and T cells from *Dh* /+ and +/+ animals failed to produce the co-operative effect which normally results in IgM synthesis (Battisto *et al.*, 1971). These experiments showed that both the T and B cell populations were defective therefore suggesting that the spleen normally exerts some influence on bone marrow and thymus derived cells. Furthermore this influence must be gestational as injection of month old *Dh* /+ mice with spleen cells from +/+ mice does not restore the immunological competence. The effect of the spleen on T cells was further demonstrated when Fletcher *et al.* (1977), again using *Dh* /+ animals, concluded that the spleen significantly effects the rate of murine T cell maturation.

Dh /+ animals have therefore been able to shed light on various immunological phenomena. Perhaps more importantly, it has been shown that *Dh* /+ mice and neo-natal splenectomised mice are quite different animal models of asplenia (Fletcher *et al.*, 1977; Welles & Battisto, 1976). Obviously splenectomised mice can only approximate hereditary asplenia as any gestational influence of the spleen on the immune system is not impaired. The gestational influence is obviously very important because transplantation of spleen cells at birth does not correct the haemotological and immunological alterations of *Dh* /+ animals (Lozzio & Machado, 1975; Machado & Lozzio, 1976).

Dh /+ animals therefore represent a good animal model for the study of the effect, particularly during gestation, of asplenia on the immune system. The *Dh* mutation has been combined with *nude* (*nu*) mice to generate the asplenic - athymic 'Lasat' strain of mice (Gershwin *et al.*, 1978).

Chapter Four

Linkage Analysis

4.1 Introduction

Homeobox-containing genes were first discovered in *Drosophila*, where they they are now known to have a vital and important function during development (Lewis, 1978; Nusslein-Volhard, 1979; Kaufman *et al.*, 1980; Nusslein-Volhard & Wieschaus, 1980). Using *Drosophila* homeobox probes cross hybridising sequences have been detected, and in several cases, shown to be developmentally expressed in a wide variety of organisms, including *Xenopus*, mouse, rat and human (Levine et al., 1984; McGinnis et al., 1984). This conservation over such a vast evolutionary span suggested that some of the genes involved in controlling mammalian development were within reach.

One murine homologue of *en*, a gene known to be involved in *Drosophila* segmentation is *En-1* (Kornberg *et al.*, 1981a, 1981b; Kornberg *et al.*, 1985; Fjose *et al.*, 1985). *En-1* was cloned and mapped, via recombinant inbred strain analysis, to the central portion of chromosome 1 (Hill *et al.*, 1987; Joyner & Martin, 1987). This finding was particularly exciting as it localised the *En-1* gene close to, or at, the previously mapped *Dh* locus.

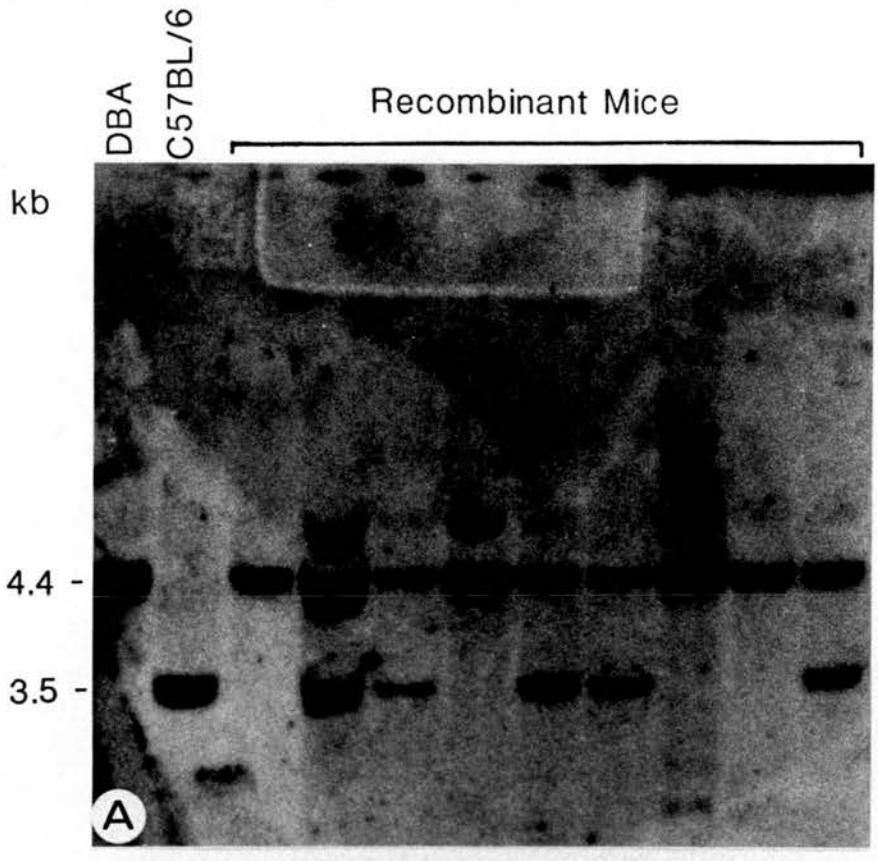
Up until this point one of the major problems with the functional analysis of mammalian homeobox-containing genes had been the lack of associated developmental mutations, which were so vital in establishing the importance and function of the *Drosophila* genes (see Nusslein-Volhard,1980). The above observation was therefore very provocative as it prompted the question of whether *Dh* represented a mutant allele of *En-1*. This chapter describes a linkage analysis which was performed, primarily to test for allelism between the two genes.

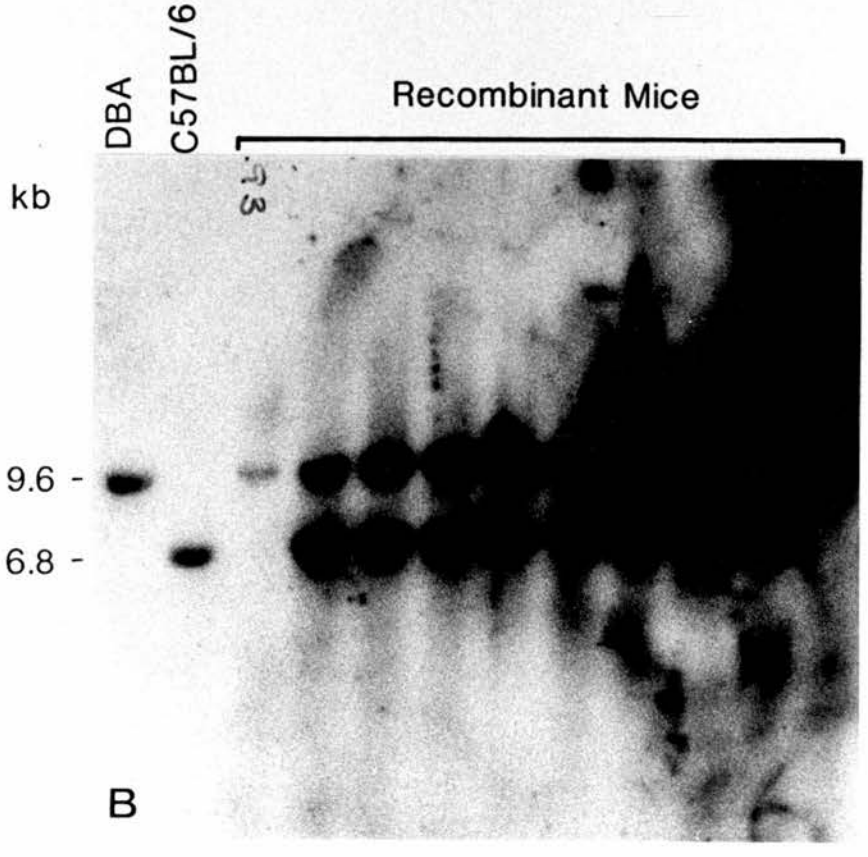
4.2 Results

A standard outcross analysis was used to determine the recombination frequency between *En-1* and *Dh* as well as three closely linked markers, the endogenous murine leukaemia virus locus *Emv-17*, leaden, (*ln*),which is a coat colour gene and the red blood cell enzyme, Peptidase-3 (*Pep-3*). The specific aim of this analysis was to determine the genetic distance and orientation of *En-1* relative to *Dh* . Heterozygous parental animals were derived from *Dh* /+ animals obtained from MRC Radiobiology Unit, Harwell.

Mice carrying *ln*, *Emv-17*^{b}, *Dh*, *En-1*^{d} and *Pep-3*^{b} in coupling were mated to inbred C57BL/Ola animals which are homozygous for +, *Emv-17*^{d}, +, *En-1*^{b} and *Pep-3*^{a} at each respective locus. The resultant F1 animals which clearly expressed *Dh* were then mated to mice from the Harwell LIII stock which were homozygous for *ln* , *Emv-17*^{b} , +, *En-1*^{d}, *Pep-3*^{b}. The *Emv* and *En* genotype was determined using restriction fragment length polymorphism (RFLP) analysis; *En-1* shows a *Taq* 1 RFLP between C57BL/6 (3.5kb) and DBA (4.4kb) mice using the pEn-1471 probe which is a 3kb *Bam*HI-*Sst*I genomic fragment containing the 3' untranslated region of the *En-1* gene (see Fig 4.1A). *Emv-17* shows a *Eco* RI RFLP using pPS1.25 which is a 1.25kb *Pst* I-*Sst* I fragment of genomic sequences flanking the 5' region of the *Emv-17* provirus of RF/J mice (Buchberg *et al.*, 1986). This probe hybridises to a 9.6kb band in DBA DNA and a 6.8kb band in C57BL/6 DNA (Fig 4.1B).The *Pep-3* genotype was determined electrophoretically (Fig 4.2) whereas *ln* and *Dh* were determined phenotypically .

Figure 4.1. Southern blot analysis of recombinant offspring showing restriction fragment length polymorphism (RFLP) for *En-1* and *EMV-17*. *A*, Genomic DNA digested with *Taq* I and hybridised with pEN-1471. *B*, Genomic DNA digested with *Eco* RI and hybridised with pPS1.25.


DBA
C57BL/6
Recombinant Mice
kb
4.4
3.5
A


DBA
C57BL/6
Recombinant Mice
kb
9.6
6.8
B

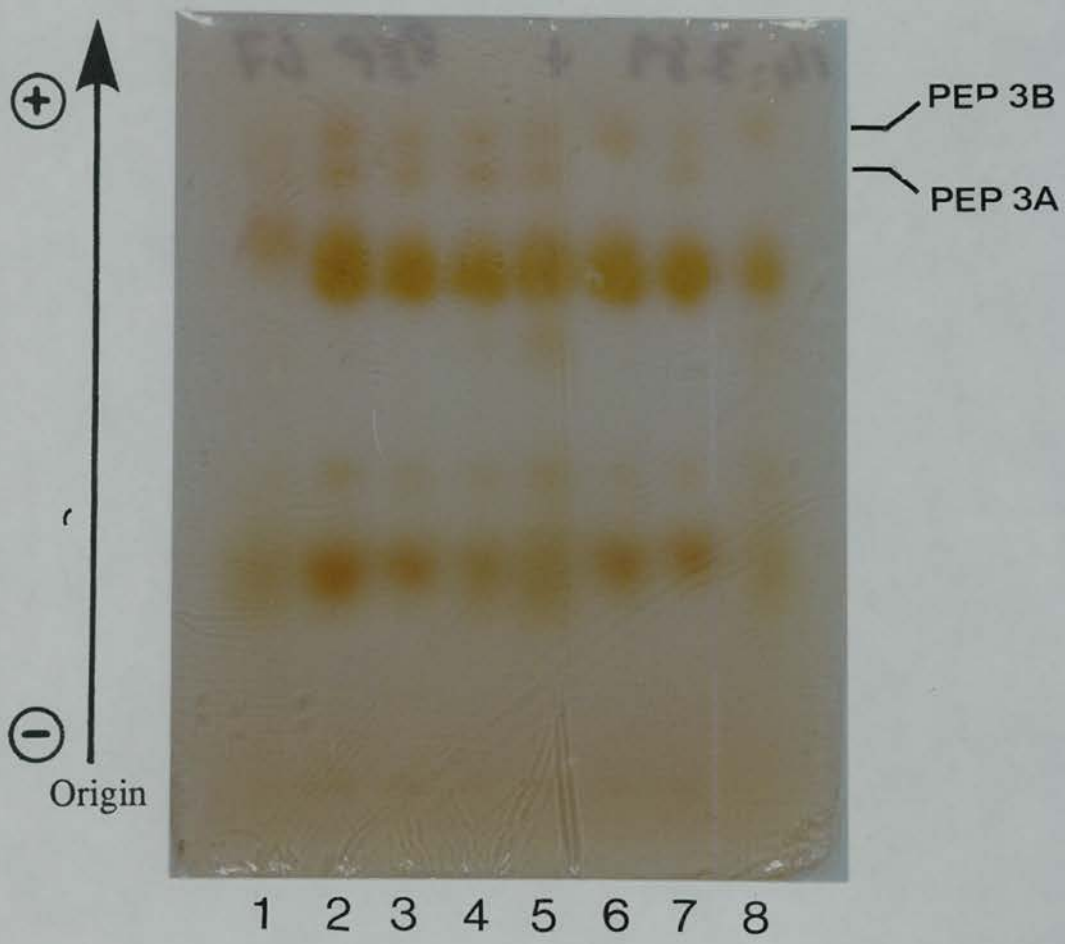


Figure 4.2 Electrophoretic pattern of *Peptidase 3* (*Pep-3*) after zip zone electrophoresis. Lanes 2-5, 7 are heterozygous for *Pep-3* ($Pep\text{-}3^{a/b}$); lanes 6 and 8 are homozygous for the DBA allele ($Pep\text{-}3^{b}$).

4.2.1 Generation and Maintenance of Stocks

1. Maintenance of Dh Stocks

FZT mice were obtained from MRC Radiobiology Unit, Harwell where they had kindly been recovered from frozen stocks by Mr. Peter Glenister. FZT mice are segregating for both *Dh* and *ln* and were maintained by random breeding of *Dh* /+ heterozygotes.

2. Generation of Partially Congenic Stock "DH"

For the purpose of this study it was decided to establish a partially congenic *Dh* strain of mice in order to standardise genetic background effects with respect to the *Dh* mutation. A *Dh* /+ mouse from the FZT stock was crossed with a C57BL/Ola mouse. The DH strain was then established by continual backcrossing of *Dh* /+ offspring onto the C57BL/Ola background (2-4 generations). The offspring which were to be used in the linkage study were assayed to ensure that the correct *Pep-3* genotype was being maintained. It was also necessary to testcross the prospective animals in order to ensure that *ln* was also being maintained. The genotype of the DH mice is given below:

ln	Emv-17^{b}	Dh	En-1^{d}	Pep-3^{b}
+	Emv-17^{d}	+	En-1^{b}	Pep-3^{a}

3. Generation of F1 Stock "DHF"

DH mice were initially used in the linkage study. However one bottleneck in the setting up of linkage matings was the testcross to ensure that the DH mice were still carrying *ln*. This problem was overcome when a *ln Dh / ln* + mouse was produced in the FZT stocks.This mouse was bred with a *ln Dh* / ++ FZT mouse to increase the numbers of *ln Dh / ln* + mice. These were then used to maintain stocks. These *ln Dh / ln* + mice were backcrossed onto the C57BL/Ola background and the F1 offspring which were *ln Emv-17^{b} Dh En-1^{d} Pep-3^{b} / + Emv-17^{d} + En-1^{b} Pep-3^{a}* were used in the linkage study. The genotype of the DHF mice at the five loci of interest is the same as that given for "DH" with the exception that they are homozygous for *ln*.

4. Generation of Genetically Defined LIII Stocks

The LIII strain of mice,which is an outbred linkage testing stock for chromosome 1, was obtained from MRC Radiobiology Unit, Harwell. LIII mice are homozygous for *ln* and wild type for *Dh*. The mice were tested for their *Pep-3*, *Emv-17* and *En-1* genotypes. Those mice which were homozygous at all three loci were used to maintain stocks.The genotype of the LIII mice is given below:

ln	Emv-17^{b}	+	En-1^{d}	Pep-3^{b}
ln	Emv-17^{b}	+	En-1^{d}	Pep-3^{b}

These mice were used as the homozygous parent in the linkage study

4.2.2 *Close Linkage of Engrailed-1 to the Dominant Hemimelia Locus on Chromosome 1*

The linkage study was carried out in two stages. Male and female mice from either the DH or the DHF stocks were mated to LIII mice and the offspring were initially scored for *ln Dh* and *Pep-3*. The results of this recombination analysis are shown in Tables 1-3. From these it can be seen that a total of 563 animals were scored of which 56 showed recombination within the *ln-Dh-Pep3* region. These 56 recombinant animals were then further analysed at the molecular level. This experimental design avoided the necessity for the molecular analysis of all offspring by focusing on those animals known to be recombinant in the *ln-Dh-Pep3* region. Tail tissue was collected from each recombinant animal for DNA preparation. Southern blots of *Taq* I digested DNA and *Eco* RI digested DNA were hybridised with the *En-1* and *Emv-17* probes respectively (see Figs 4.1a, 4.1b). The results of this further analysis are shown in Tables 4-6. From this data it can be seen that although *Dh* and *En-1* are closely linked they are separable by recombination (4/563). The likely gene order and recombination frequencies of these loci are:

ln (5.2 ± 0.9) *Emv-17* (1.1 ± 0.4) *Dh* (0.7 ± 0.4) *En-1* (3.0 ± 0.7) *Pep-3*

Table 1. Recombination Events Observed Between the ln-Dh-Pep-3 Loci in the Cross DH Female x LIII Male or DH Male x LIII Female.

Cross:	ln	Dh	Pep-3^{b}	ln	+	Pep-3^{b}
	+	+	Pep-3^{a}	ln	+	Pep-3^{b}

Gene Combinations	ln	Dh	Pep-3	Heterozygous Parent		
				Female	Male	Total
Parental Type	ln	Dh	b	12	60	72
	+	+	a	34	156	190
Recombinant (ln-Dh)	ln	+	a	2	6	8
	+	Dh	b	1	5	6
Recombinant (Dh-Pep-3)	ln	Dh	a	0	0	0
	+	+	b	3	3	6
Double Recombinants	ln	+	b	0	0	0
	+	Dh	a	0	0	0

Recombinant Fractions:	Heterozygous Female	Heterozygous Male	Total
ln-Dh	3/52 = 5.8± 3.2	11/230 = 4.8 ± 1.4	14/282 = 5.0 ± 1.3
Dh-Pep-3	3/52 = 5.8 ±3.2	3/230 = 1.3 ± 0.7	6/282 = 2.1 ± 0.9
ln-Pep-3	6/52 = 11.5 ± 4.4	14/230 = 6.1 ±1.6	20/282 = 7.1 ± 1.5

Table 2 Recombination Events Observed Between ln-Dh-Pep-3 Loci in the Cross DHF Female x LIII Male or DHF Male x LIII Female.

Cross:	ln	Dh	Pep-3^b	ln	+	Pep-3^b
	+	+	Pep-3^a	ln	+	Pep-3^b

Gene Combinations	ln	Dh	Pep-3	Heterozygous Parent		
				Female	Male	Total
Parental Type	ln	Dh	b	61	32	93
	+	+	a	90	62	152
Recombinant (ln-Dh)	ln	+	a	16	1	17
	+	Dh	b	3	1	4
Recombinant (Dh-Pep-3)	ln	Dh	a	5	0	5
	+	+	b	7	3	10
Double Recombinants	ln	+	b	0	0	0
	+	Dh	a	0	0	0

Recombinant Fractions:	Heterozygous Female	Heterozygous Male	Total
ln-Dh	19/182 = 10.4 ± 2.3	2/99 = 2.0 ± 1.4	21/281 = 7.5 ± 1.6
Dh-Pep-3	12/182 = 6.6 ± 1.8	3/99 = 3.0 ± 1.7	15/281 = 5.3 ± 1.3
ln-Pep-3	31/182 = 17.0 ± 2.8	5/99 = 5.1 ± 2.2	36/281 = 12.8 ± 2.0

Table 3 Total Recombination Events Observed Between the ln-Dh-Pep-3 Loci in Both the DH and DHF Crosses.

Cross:	ln	Dh	Pep-3^{b}	ln	+	Pep-3^{b}
	+	+	Pep-3^{a}	ln	+	Pep-3^{b}

Gene Combinations	ln	Dh	Pep-3	Heterozygous Parent		
				Female	Male	Total
Parental Type	ln	Dh	b	73	92	165
	+	+	a	124	218	342
Recombinant (ln-Dh)	ln	+	a	18	7	25
	+	Dh	b	4	6	10
Recombinant (Dh-Pep-3)	ln	Dh	a	5	0	5
	+	+	b	10	6	16
Double Recombinants	ln	+	b	0	0	0
	+	Dh	a	0	0	0

Recombinant Fractions:	Heterozygous Female	Heterozygous Male	Total
ln-Dh	22/234 = 9.4 ± 1.9	13/329 = 4.0 ± 1.1	35/563 = 6.2 ± 1.0
Dh-Pep-3	15/234 = 6.4 ± 1.6	6/329 = 1.8 ± 0.7	21/563 = 3.7 ± 0.8
ln-Pep-3	37/234 = 15.8 ± 2.4	19/329 = 5.8 ±1.3	56/563 = 9.9 ± 1.3

Table 4 Molecular Analysis of the Recombinant Animals Observed in the Cross DH Female x LIII Male or DH Male x LIII Female.

Cross: ln Emv-17^{b} Dh En-1^{d} Pep-3^{b} ln Emv-17^{b} + En-1^{d} Pep-3^{b}

+ Emv-17^{d} + En-1^{b} Pep-3^{a} ln Emv-17^{b} + En-1^{d} Pep-3^{b}

Gene combinations:					Heterozygous Parent		
ln	Emv-17	Dh	En-1	Pep-3	Female	Male	Total
ln-Dh Recombinants:							
ln	d	+	b	a	1	5	6
+	b	Dh	d	b	1	4	5
ln	b	+	b	a	1	1	2
+	d	Dh	d	b	0	1	1
Dh-Pep-3 Recombinants:							
ln	b	Dh	b	a	0	0	0
+	d	+	d	b	0	1	1
ln	b	Dh	d	a	0	0	0
+	d	+	b	b	3	2	5

No other gene combinations were found.

Recombinant Fractions:	Heterozygous Female	Heterozygous Male	Total
ln - Emv-17	2/52 = 3.8 ± 2.7	9/230 = 3.9 ± 1.3	11/282 = 3.9 ± 1.2
Emv-17 - Dh	1/52 = 1.9 ± 1.9	2/230 = 0.9 ± 0.6	3/282 = 1.1 ± 0.6
Dh - En-1	0/52 = 0.0 ± 0.0	1/230 = 0.4 ± 0.4	1/282 = 0.4 ± 0.4
En-1 Pep-3	3/52 = 5.8 ± 3.2	2/230 = 0.9 ± 0.6	5/282 = 1.8 ± 0.8
ln - En-1	3/52 = 5.8 ± 3.2	12/230 = 5.2 ± 1.5	15/282 = 5.3 ± 1.3
Emv-17 - En-1	1/52 = 1.9 ± 1.9	3/230 = 1.3 ± 0.7	4/282 = 1.4 ± 0.7
Emv-17 Pep-3	4/52 = 7.7 ± 3.7	5/230 = 2.2 ± 1.0	9/282 = 3.2 ± 1.0

Table 5 Molecular Analysis of the Recombinant Animals Observed in the Cross DHF Female x LIII Male or DHF Male x LIII Female.

Cross: ln Emv-17^{b} Dh En-1^{d} Pep-3^{b} ln Emv-17^{b} + En-1^{d} Pep-3^{b}

+ Emv-17^{d} + En-1^{b} Pep-3^{a} ln Emv-17^{b} + En-1^{d} Pep-3^{b}

Gene combinations:					Heterozygous Parent		
ln	Emv-17	Dh	En-1	Pep-3	Female	Male	Total
ln-Dh Recombinants:							
ln	d	+	b	a	14	1	15
+	b	Dh	d	b	2	1	3
ln	b	+	b	a	2	0	2
+	d	Dh	d	b	1	0	1
Dh-Pep-3 Recombinants:							
ln	b	Dh	b	a	1	0	1
+	d	+	d	b	1	1	2
ln	b	Dh	d	a	4	0	4
+	d	+	b	b	6	2	8

No other gene combinations were found.

Recombinant Fractions:	Heterozygous Female	Heterozygous Male	Total
ln - Emv-17	16/182 = 8.8 ± 2.1	2/99 = 2.0 ± 1.4	18/281 = 6.4 ± 1.5
Emv-17 - Dh	3/182 = 1.6 ± 0.9	0/99 = 0.0 ± 0.0	3/281 = 1.1 ± 0.6
Dh - En-1	2/182 = 1.1 ± 0.8	1/99 = 1.0 ± 1.0	3/281 = 1.1 ± 0.6
En-1 Pep-3	10/182 = 5.5 ± 1.7	2/99 = 2.0 ± 1.4	12/281 = 4.3 ± 1.2
ln - En-1	21/182 = 11.5 ± 2.4	3/99 = 3.0 ± 1.7	24/281 = 8.5 ± 1.7
Emv-17 - En-1	5/182 = 2.7 ± 1.2	1/99 = 1.0 ± 1.0	6/281 = 2.1 ± 0.9
Emv-17 Pep-3	15/182 = 8.2 ± 2.0	3/99 = 3.0 ± 1.7	18/281 = 6.4 ± 1.5

Table 6 Molecular Analysis of All Recombinant Animals Observed in Both the DH and DHF Crosses.

Cross: ln Emv-17^{b} Dh En-1^{d} Pep-3^{b} ln Emv-17^{b} + En-1^{d} Pep-3^{b}

+ Emv-17^{d} + En-1^{b} Pep-3^{a} ln Emv-17^{b} + En-1^{d} Pep-3^{b}

Gene combinations:					Heterozygous Parent		
ln	Emv-17	Dh	En-1	Pep-3	Female	Male	Total
ln-Dh Recombinants:							
ln	d	+	b	a	15	6	21
+	b	Dh	d	b	3	5	8
ln	b	+	b	a	3	1	4
+	d	Dh	d	b	1	1	2
Dh-Pep-3 Recombinants:							
ln	b	Dh	b	a	1	0	1
+	d	+	d	b	1	2	3
ln	b	Dh	d	a	4	0	4
+	d	+	b	b	9	4	13

No other gene combinations were found.

Recombinant Fractions:	Heterozygous Female	Heterozygous Male	Total
ln - Emv-17	18/234 = 7.7 ± 1.7	11/329 = 3.3 ± 1.0	29/563 = 5.2 ± 0.9
Emv-17 - Dh	4/234 = 1.7 ± 0.8	2/329 = 0.6 ± 0.4	6/563 = 1.1 ± 0.4
Dh - En-1	2/234 = 0.9 ± 0.6	2/329 = 0.6 ± 0.4	4/563 = 0.7 ± 0.4
En-1 Pep-3	13/234 = 5.6 ± 1.5	4/329 = 1.2 ± 0.6	17/563 = 3.0 ± 0.7
ln - En-1	24/234 = 10.3 ± 2.0	15/329 = 4.6 ± 1.2	39/563 = 6.9 ± 1.1
Emv-17 - En-1	6/234 = 2.6 ± 1.0	4/329 = 1.2 ± 0.6	10/563 = 1.8 ± 0.6
Emv-17 Pep-3	19/234 = 8.1 ± 1.8	8/329 = 2.4 ± 0.8	27/563 = 4.8 ± 0.9

4.2.3 Factors Affecting Recombination Frequency

As two strains of heterozygous parent were used in this study the data was analysed to see if any significant differences in recombination frequency occurred either between or within the two crosses. The data are given below.

Table 7. Observed % Recombination Over the Entire In-Dh-Pep-3 Region in Both the DH and the DHF Cross.

DH Cross	Offspring			% Recombination
Heterozygous Parent	Parental	Recombinant	Total	
Female	46	6	52	11.54 ± 4.43
Male	216	14	230	6.09 ± 1.58
Total	262	20	282	7.09 ± 1.53

DHF Cross	Offspring			% Recombination
Heterozygous Parent	Parental	Recombinant	Total	
Female	151	31	182	17.03 ± 2.79
Male	94	5	99	5.05 ± 2.20
Total	245	36	281	12.81 ± 1.99

Chi square tests were performed in order to determine whether any significant differences were found in recombination frequencies (R.F.) either between or within each cross. When comparing the r.f. between either DH female : DHF female (χ^2 = 0.55, P>0.05) or DH male : DHF male (χ^2 = 0.01, P> 0.05) no significant

differences were found. However when comparing within each cross significant differences can be seen ; DH female : DH male $\chi^2 = 1.17$, $P < 0.05$ and DHF female : DHF male gives a χ^2 value of 7.20, $P<0.01$. In both cases the heterozygous female parent has a higher recombination frequency.

Although intercross comparisions of either female : female or male : male show no significant differences when one considers each cross as a whole ie DH male and female compared to DHF male and female there is a significant difference ($\chi^2 = 4.52$, $P<0.05$). The DHF cross exihibts a higher recombination frequency than does the DH cross (see discussion below).

4.3 Discussion

A major goal in mammalian genetics is the construction of a high resolution linkage map which can provide the basis for the construction of physical maps. Physical maps can then be directly applied to reverse genetic approaches for the isolation of various genes such as *Dh*. For this to be achieved it is necessary that observed recombination frequencies are as accurate as possible. In this study two different stocks of heterozygous parent were used and it is essential that in order to obtain an accurate map any differences in recombination frequency between the two strains must be investigated

If one considers the recombination frequency between the crosses in relation to the sex of the heterozygous parent i.e. comparing DH female with DHF female or DH male with DHF male no significant differences are detectable. This is not suprising as the relevant region in the heterozygous parent in either cross is essentially the same. However when one compares the recombination frequency obtained between sexes,within each cross, significant differences are detected. In either cross the female exhibits a higher recombination frequency than its male counterpart and this difference is more significant in the DHF cross. That there should be such a difference in recombination frequency between males and females is not to suprising as a large number of researchers have reported a similar finding (Lyon & Searle, 1989). Indeed in mice the recombinational map in females is some 13% longer than that of males (Roderick and Hillyard, 1989). Why females should exhibit a higher recombination frequency than males has not yet been established but it may be a reflection the meiotic differences of oogenesis and spermatogenesis. A consideration of the overall recombination frequency obtained for each cross shows that the two are significantly different. The DH cross has an overall recombination frequency of 7.09% whereas the DHF cross exhibits an overall recombination frequency of 12.8%. As described above

intercross comparisions related to the sex of the heterozygous parent shows no obvious differences. Therefore the difference seen in the overall recombination frequency is probably a reflection of the different ratios, of male to female heterozygous parent, used in either cross. Of fifteen DH matings only two (13%) had a female heterozygous parent whereas of the ten DHF matings six (60%) had a female heterozygous parent. The corresponding proportions of offspring from female heterozygotes is 52/282 (18.4%) for DH and 182/281 (64.8%) for DHF matings.

The results presented here show that although *Dh* and *En-1* are closely linked together on the same region of chromosome 1, they are separable by recombination and *Dh* maps 0.7cM proximal to *En-1* and 1.1cM distal to *Emv-17* .These data provides genetic evidence that *Dh* is unlikely to simply be a mutant allele of *En-1*. This result is substantiated by Martin *et al.* (1990) who were also able to demonstrate recombination between *Dh* and *En-1*. However one major discrepancy between the two studies is the reported genetic map distances. Martin *et al* mapped *Dh* to a position 0.28cM ± 0.28cM proximal to *En-1* and 0.28cM ± 0.28cM distal to *Emv-17* , thus rendering a genetic distance of 0.55cM over the entire *Emv-17* - *En-1* region compared to 1.8cM estimated in this study. This difference could be due to (1) differences in genetic background between the two studies, (2) different proportions of offspring from male and female *Dh* /+ heterozygotes or (3) chance.

In the mouse the penetrance and expressivity of mutations can be influenced by the genetic background on which the mutation is present eg a variation in pleiotropic effects can be observed when the mutation *shortear* (*se*) is put onto different backgrounds (Green, 1957). Furthermore some dominant mutations exhibit a parental origin specific difference when put onto different genetic backgrounds eg the penetrance of the maternally inherited *Fused* mutation can be decreased on certain backgrounds whereas the penetrance of paternally inherited *Fused* remains high

(Agulnik & Ruvinsky, 1988) However the recovery of *Dh* /+ animals from Martin's study was 44.5% (classification occured at birth) which is in accordance with Searle's work Searle, 1964; see chapter three). If misclassification of *Dh* /+ animals occurred due to lack of penetrance / expressivity Martin *et al.* would have reported double recombinants. It therefore seems unlikely that the discrepancy between the estimated recombination frequencies can be accounted for by differences in genetic background.

The discrepancy between the two studies is also unlikely to be the result of differences between male and female recombination frequencies. As mentioned previously females tend to exhibit a higher recombination frequency than do males. If the study by Martin *et al.* had only used male heterozygous parents the recombination frequency might be substantially lower. This however is not the case as both studies present data which is composed from both male and female heterozygous parents. Thus it would seem more plausible to suggest that the differences in recombination frequencies are due to chance and sample size. The data presented here involved the analysis of 563 offspring compared to the 362 offspring from Martin's data.

The data presented in this thesis estimates a genetic map distance of 1.8cM between *Emv-17* and *En-1* which is in good agreement with previously published data (Buchberg *et al.*, 1986) but contrasts greatly with the map distance reported by Martin *et al.* From Fig 4.3 it can be seen that, although it places *ln* slightly more proximal, the map presented here is, overall, in good agreement with published data as given in Mouse Genome, Vol 87 (1990). Although the genetic map presented by Martin *et al.* also orientates *Dh* proximal to *En-1*, it is inconsistent with the published data. It is therefore likely that the map distance presented here, which places *Dh* 0.7cM proximal to *En-1*, is a more accurate estimate. However as the discrepancies between the two studies are most probably a reflection of sample size combining the data may in fact give a more accurate genetic map distance. Thus combining the data with regard

to recombination between *Dh* and *En-1* (5/925) gives a recombination frequency of 0.54 ± 0.2. This genetic distance of 0.54cM shall be used as the estimated genetic distance of *Dh* relative to *En-1* in the discussion below.

As *Dh* and *En-1* are separable by recombination this would argue that *Dh* is not simply a mutant allele of *En-1*. Indeed *En-1* mRNA transcripts are of normal size and abundance in +/+, *Dh* /+ and *Dh/Dh* embryos (Martin *et al.*, 1990). This indicates that, certainly at a gross level, *En-1* is not responsible for the *Dh* mutation. This however does not rule out the possibility that *Dh* and *En-1* do have some developmental relationship and that the *Dh* gene causes anomolous expression of *En-1* in *Dh* /+ and *Dh/Dh* animals. En-1 certainly does show expression in several tissues affected by *Dh* (vertebrae, ribs and limb buds) as well as tissues apparently unaffected by *Dh* (CNS) (Davidson *et al.*, 1988; Davis & Joyner, 1988; Davis *et al.*, 1991). It does not however show any signs of expression in the splanchnic mesoderm which Green (1967) proposed was the primary site of action of the *Dh* gene. If however *Dh* is a mutation which primarily affects segmentation as postulated in the introduction to this thesis it is certainly interesting to note that *En-1* is expressed periodically throughout the length of the pericordal tube (Davis & Joyner, 1988). The pericordal tube is established when a subset of sclerotomal cells migrate ventrally and axially to surround the notochord (Verbout, 1985). The expression of *En-1* within the loosely packed cells of the pericordal tube, and which later gives rise to the vertebrae appears to be the exact complement of *Pax-1* expression, which is only expressed in the tightly packed cells of the pericordal tube (Deustch *et al.*, 1988). As both genes are expressed at more or less the same time in development it has been suggested that *En-1* and *Pax-1* are involved in the coordination of the development of the vertebrae. If *En-1* does have a role in vertebrae formation then it may be that the *Dh* gene interferes with the normal role of *En-1* during this aspect of development.

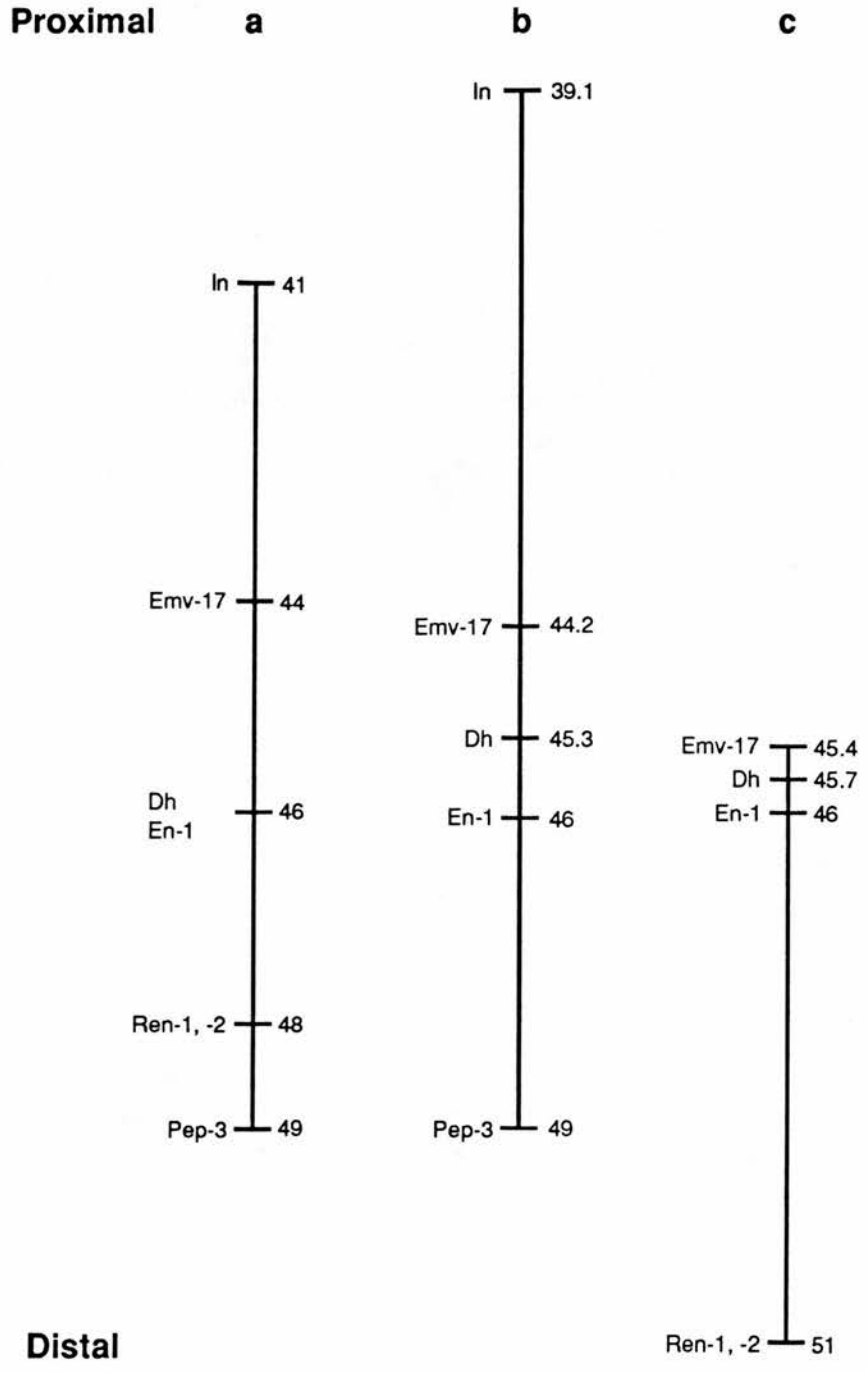


Figure 4.3. Comparision of genetic maps. (a), map from Mouse Genome vol.87, 1990. (b), map from data presented in this thesis. (c), map from data presented by Martin *et al.*(1990). Maps are centred on *En-1*.

The genetic map of the mouse is approximately 1600cM and the haploid genome content is approximately 3 x 10^9 base pairs (bp). If the relationship between physical and genetic distance is constant one would expect that 1cM would equal approximately 2Mbp. The genetic distance of 0.54 cM between *Dh* and *En-1* would therefore be equivalent to a physical distance of approximately 1.1Mb. However the mouse genome does not adhere to this average (1cM = 2Mb) in a uniform way. Indeed in certain regions of the mouse genome 1cM is equivalent to 4 Mb (L. Stubbs, pers. comm.) and a comparision of the physical and genetic maps around the *agouti* locus suggests that the ratio of physical to genetic distance in different regions exhibits a wide variation . In two adjacent segments of the *agouti* locus Barsh & Epstein (1989) reported values of <150kb/cM and 4Mb/cM. In accordance with this data *Dh* could be as much as 2.16Mb or as little as 80kb away from *En-1* in molecular terms.

In the mouse there are two *En* genes, *En-1* and *En-2* , which share extensive homology both with each other and with their Drosophila counterpart (Joyner *et al.*, 1985; Joyner & Martin, 1987). We have just shown that *En-1* maps 0.7cM distal to *Dh* on chromosome 1 and Martin *et al.* (1990) reported that *En-2* maps 1.1cM proximal to hemimelic extra toes (*Hx*) on chromosome 5. *Dh* and *Hx* both cause similar phenotypic abnormalities :- *Dh* causes skeletal abnormalities of the hind limb which include preaxial poly- and oligodactly and a reduction or absence of the tibia. *Hx* is a dominant mutation which causes preaxial polydactly and hemimelia (Dickie, 1968; Knudsen & Kochhar, 1981). That each of the *En* genes show genetic linkage with essentially similar developmental mutations is certainly very provocative. Although the significance of these linkages has not been ascertained it is inviting to suggest that it is more than coincidence. As the vertebrate genome is believed to have undergone several duplications during evolution (Nadeau, 1989) these linkages may represent a duplicated conserved linkage group. It has been suggested that the

ecotropic proviral integration locus *Emv-17* is also part of this linkage group (Martin *et al.*, 1990). *Emv-1,* which is also an ecotropic proviral integration locus, is closely linked to *En-2* and *Hx* on chromosome 5 (Martin *et al.*, 1990). This would suggest that a common ecotropic proviral integration site existed in the ancestral genome segment. The actual proviral sequences would not be part of this group as their insertion is a relatively recent event, since they are found in some strains of laboratory mice and not others (Jenkins *et al.*, 1982). Although ecotropic proviruses are found all over the genome there is evidence to suggest that there are preferential sites for proviral integration near developmentally active genes (Shih *et al.*, 1983). If this region of the genome does represent a conserved linkage group it would suggest that *Dh* and *Hx,Hm* are paralogous genes. The molecular cloning of the *Dh* gene may help facilitate the cloning of the *Hx* gene and the tightly linked gene Hammertoe (*Hm;* Sweet, 1982).

Chapter Five

Long and Short Range Physical Mapping around the *En-1* Locus

5.1 Introduction

The isolation of loci known only from their mutant phenotype first requires the genetic mapping of the locus followed by the application of molecular techniques to identify and isolate coding sequences that may represent the mutant locus.

At the beginning of this project two mouse cDNA clones, one which represented *En-1* and the other *En-2* , had been isolated via the *Drosophila engrailed* cDNA probe from a 8.5day mouse embryo cDNA library. The *En-1* cDNA, pλ4, was subsequently mapped, using recombinant inbred strains, to chromosome 1 near or at the developmental locus *Dh*. Initial molecular studies with pλ4 involved screening for any abnormalities of *En-1* in *Dh* /+ mice. From this analysis it was concluded that if *En-1* was responsible for the *Dh* mutation then a deletion / insertion undetectable on conventional Southern blots, or a point mutation was responsible (Hill *et al.*, 1987). Subsequent work, as described in the previous chapter, found that *Dh* maps 0.7cM proximal to *En-1* and is therefore unlikely to be a mutant allele of *En-1*.

As we now know that these two genes are not alleles of one another we need to begin looking for neighbouring genes which may represent the *Dh* locus. As the orientation of *En-1* relative to the chromosome was unknown we need to extend this search in both directions. Pulsed field gel electrophoresis (PFGE) was first described in 1984 by Schwartz + Cantor and, due to its ability to resolve DNA fragments up to 10 megabases, greatly facilitates physical gene mapping and the analysis of gene rearrangements such as translocations, inversions and large deletions and insertions (reviewed by Barlow & Lehrach, 1987). Within the context of this project PFGE was employed as a means to (1) establish a bidirectional long range map around *En-1*, (2) orientate the *En-1* gene relative to *Dh*, and (3) look for chromosomal rearrangements in *Dh/Dh* DNA.

The ability to construct long range maps, certainly in terms of gene mapping, is of enormous value and is made possible due to existence of several "rare-cutting" enzymes. These enzymes contain the dinucleotide CpG in their recognition sequence and are "rare-cutters" for several reasons including the under representation of this dinucleotide in the genome (Razin & Riggs, 1980) and the inhibition of cleavage due to methylation of the cytosine residue in the enzyme recognition site (Brown & Bird, 1986). An added advantage of these rare cutting enzymes is that many of them have recognition sites within "CpG" islands (Bird, 1987; W. Bickmore, unpublished results). CpG islands are clusters of undermethylated CpG dinucleotides associated with the 5' end of genes (Bird, 1986). The ability to determine the presence and position of CpG islands, and therefore by inference transcriptionally active genes, is a great advantage particularly when one considers that the vertebrate genome has vastly more DNA than required for the specification of proteins (Alberts *et al.*, 1983).

When constructing a long range map one needs to be able to orientate the map and, as is required here, extend it in both directions. In order to establish a bidirectional long range map one must have a reference point for the enzyme sites together with flanking probes.This chapter presents both a short range and a long range map around *En-1*.The short range map was achieved via the isolation and characterisation of two cosmid clones. The mapping of these cosmids was vital for two main reasons. Firstly, it located many rare cutting enzyme sites which were found to be clustered within the *En-1* gene. Secondly, it generated probes which hybridised specifically to sequences flanking this cluster of sites. The generation of these probes allowed the long range map to extend in both directions from the *En-1* gene.

The resulting long range map, which was established via PFGE, has several important uses. Firstly, it will localise any CpG islands and therefore, by inference, genes. Although this is in itself is important, the question remains of how one will be

able to specifically identify the *Dh* gene. One way to pinpoint the locus would be to locate the recombination breakpoints of the four animals recovered from the linkage analysis, which are recombinant between *En-1* and *Dh*. Location of these breakpoints will provide data as to the minimum distance of *Dh* relative to *En-1*. Another advantage of PFGE is that it allows one to locate distant chromosomal rearrangements such as deletions and insertions which might affect the Dh gene. Finally a long range map will allow the orientation of YAC clones covering the region (see chapter six).

5.2 Results

5.2.1 *Screening of a cDNA Library*

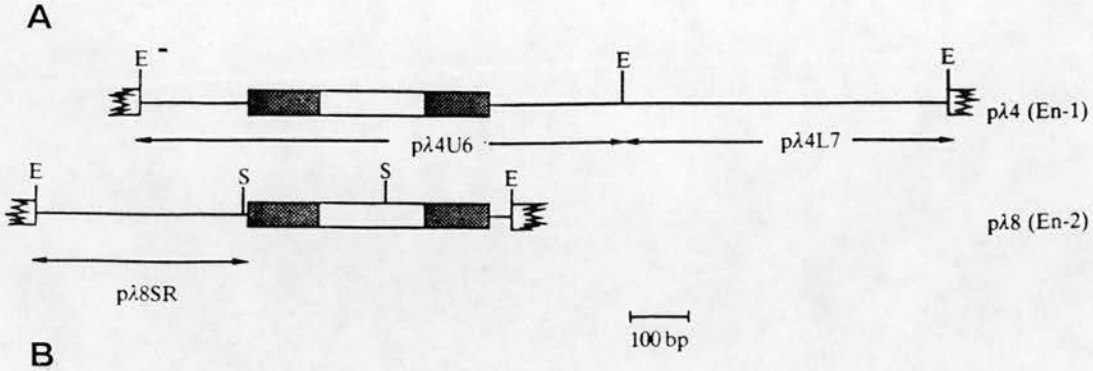
Using *Drosophila engrailed* probes two cDNA clones were recovered from a 8.5 day mouse embryonic library (Hill *et al.*, 1987). One of the clones represented *En-1* (pλ4) and the other clone represented *En-2* (pλ8). pλ4 can be restricted by *Eco* RI to give two fragments - pλ4U6 and pλ4L7 (see Fig 5.1a). pλ4U6 contains the homeobox region and therefore cross hybridises with *En-2*, whereas pλ4L7 contains the 3' untranslated region and is *En-1* specific. Both probes detect a *Taq* I polymorphism between the mouse strains DBA and C57BL/6 DNA. pλ4U6 detects a common band of 1.5 kb (*En-2*) and a polymorphic band which migrates at 1.4 kb in C57BL/6 and 4.2 kb in DBA, whereas pλ4L7 detects a polymorphic band which migrates at 3.5kb in C57BL/6 and 4.4kb in DBA (see Fig.5.1b)

As pλ4L7 is *En-1* specific this was the probe of choice in PFGE mapping and the screening of the cosmid library. pλ4L7 was also used to screen a λgt10 genomic library. This screen produced two clones, En.1S and En.1L, which were provisionally mapped (data not shown).

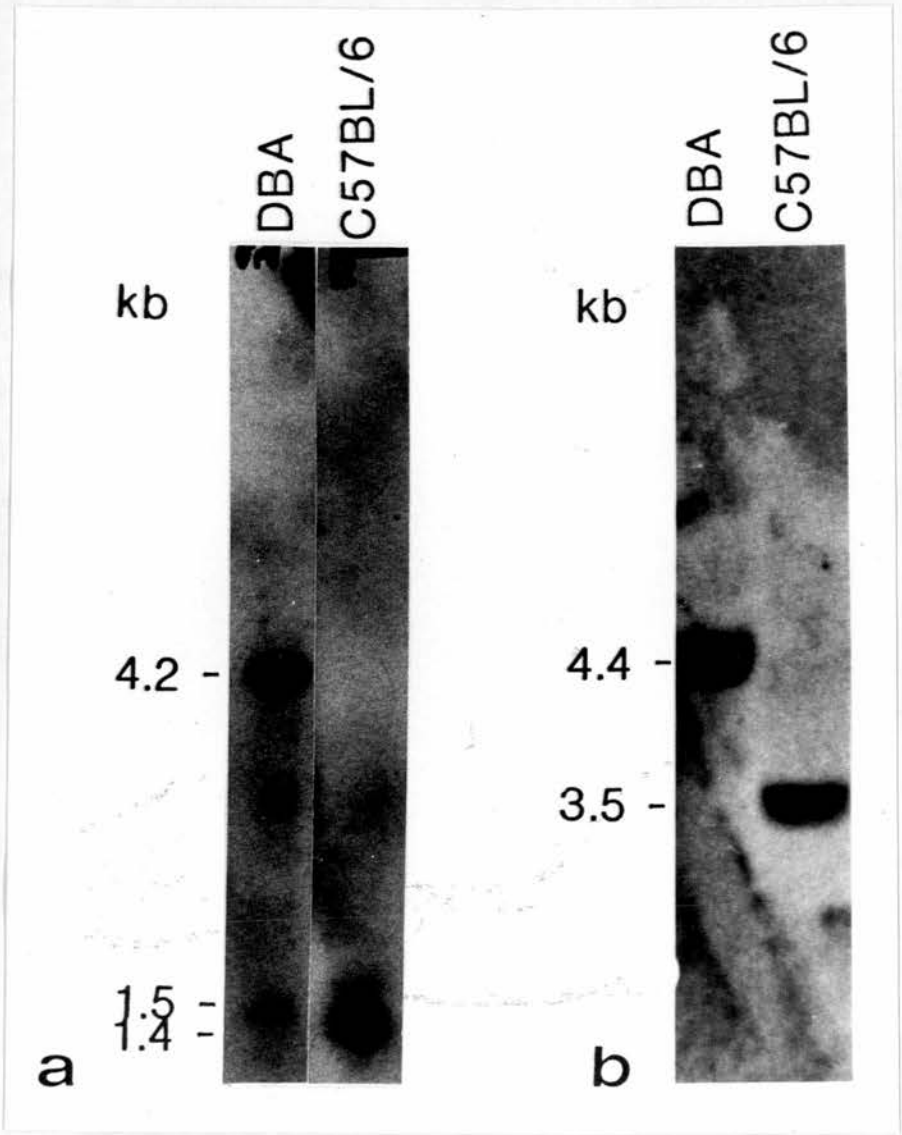
5.2.2 *Screening of a Cosmid Library*

A cosmid library was obtained from Dr L Stubbs (ICRF, London). The library was prepared by *Sau* 3A partial digestion of 129/Sv spleen DNA and cloned into the *Bam* HI site of the tetracyclin gene within the cosmid vector pcos2EMBL (Ehrich *et al.*, 1987). Five independently packaged aliquots were supplied, each containing about one genome equivalent of mouse DNA.

Figure 5.1.(*A*) *En-1* and *En-2* cDNAs and probes. Sequence similarity between the two genes includes the homeobox (open boxes) and 100 bp to the 5' side and 96bp to the 3' side (shaded boxes). The probes pλ4U6 and pλ4L7 are indicated (reproduced with permission from Davidson *et al.*, 1988). (*B*) Southern blot analysis of mouse strains DBA and C57Bl/6. Genomic DNA was digested with *Taq* I and hybridised with (a) pλ4U6 or (b) pλ4L7.


A
E
E
E
pλ4 (En-1)
pλ4U6
pλ4L7
E
S
S
E
pλ8 (En-2)
pλ8SR
100 bp

B


DBA
C57BL/6
kb
4.2
1.5
1.4
a
DBA
C57BL/6
kb
4.4
3.5
b

Each aliquot of the library was screened by plating out 200,000 colony forming units onto a 20x20 cm^2 Hybond-N membrane. After 12 hours growth, two replica filters were made from this master plate and all three plates were left to recover for 3-4 hours at 37^0C. The filters were then processed using the Grunstein & Hogness protocol. The replica filters were screened with the λ4L7 (*En-1* specific) probe. Duplicate positive colonies were picked and subjected to secondary and tertiary screening after which they could be individually isolated. The screen with λ4L7 produced six cosmid colonies. These 6 cosmids were grown up as mini preps for further analysis. DNA from all 6 were digested with *Bam* HI and electrophoresed. From the restriction fragments generated it was clear that cosmids 1,2,3,5 and 6 were the same whereas cosmid 4 was different. Cosmids 3 and 4 were named pCB3 and pCB4 and grown up for large scale DNA preps.

5.2.3 *Cosmid Mapping*

Cosmids can contain up to 40kb of insert DNA. This makes mapping by complete digestion extremely difficult if not impossible since most restriction enzymes with a 6 bp recognition site will cut within the insert several times Mapping strategies using partial digestion have been described but these require a linear starting molecule (Rackwitz *et* al., 1985). The λ phage encoded terminase can be used to both linearise cosmids at a unique sequence which is unlikely to occur randomly in any cloned DNA, and provides the single strand ends to which radiolabelled oligonucleotides can hybridise (Rackwitz *et al.*, 1985). A bacterial strain which overproduced the terminase was provided with the library, however the crude enzyme extract produced by following the method of Rackwitz resulted not only in the linearisation of the cosmids but also in their extensive degradation (Dr J. Inglis, pers. comm.) As no commercial purified terminase was available at this stage the following strategy was used.

Cosmids were first digested with *Sal* I which has a single recognition site in pcos2EMBL, within the tetracyclin resistance gene. pCB4 did not contain a *Sal* I site within the insert. pCB3 did however contain an internal *Sal* I site. This site was initally mapped to approximately the mid-point of the insert and would therefore not be a problem with regard to the mapping strategy. As well as a linear starting molecule it is also necessary to have a specific DNA probe for each end of the molecule. These were obtained by isolating and subcloning the *Bam* HI - *Sal* I (BS probe) and the *Sal* I - *Acc* I (SA probe) sequences flanking the *Sal* I site of pBR322 which is the donor plasmid for the tetracyclin gene.

Complete and partial digests were performed on aliquots of *Sal* I digested pCB3 and pCB4. Partial digests were performed with *Pst* I, *Bam* HI, *Hind* III and *Eco* RI, and were performed by diluting the enzyme to a concentration of one unit of enzyme per microlitre in 1x restriction buffer, followed by digestion of approximately 1μg of DNA with 1 unit of enzyme for 0', 2', 5', 10' and to completion. Complete and partial digests with other enzymes, in particular rare-cutting enzymes such as *Bss* HII and *Not* I were also performed. All cosmid digests were run on 0.4% agarose gels and run at 50 volts overnight. The gels were then photographed and Southern blotted onto Hybond-N membranes. The filter was then sequentially probed with the BS or the SA probes. The resulting autoradiographs gave a pattern of restriction site positions, with respect to the molecule ends, for each enzyme (see Fig.5.2). These were then sized using the Gelsize program and a restriction map of each cosmid constructed (Figs.5.3, 5.4).

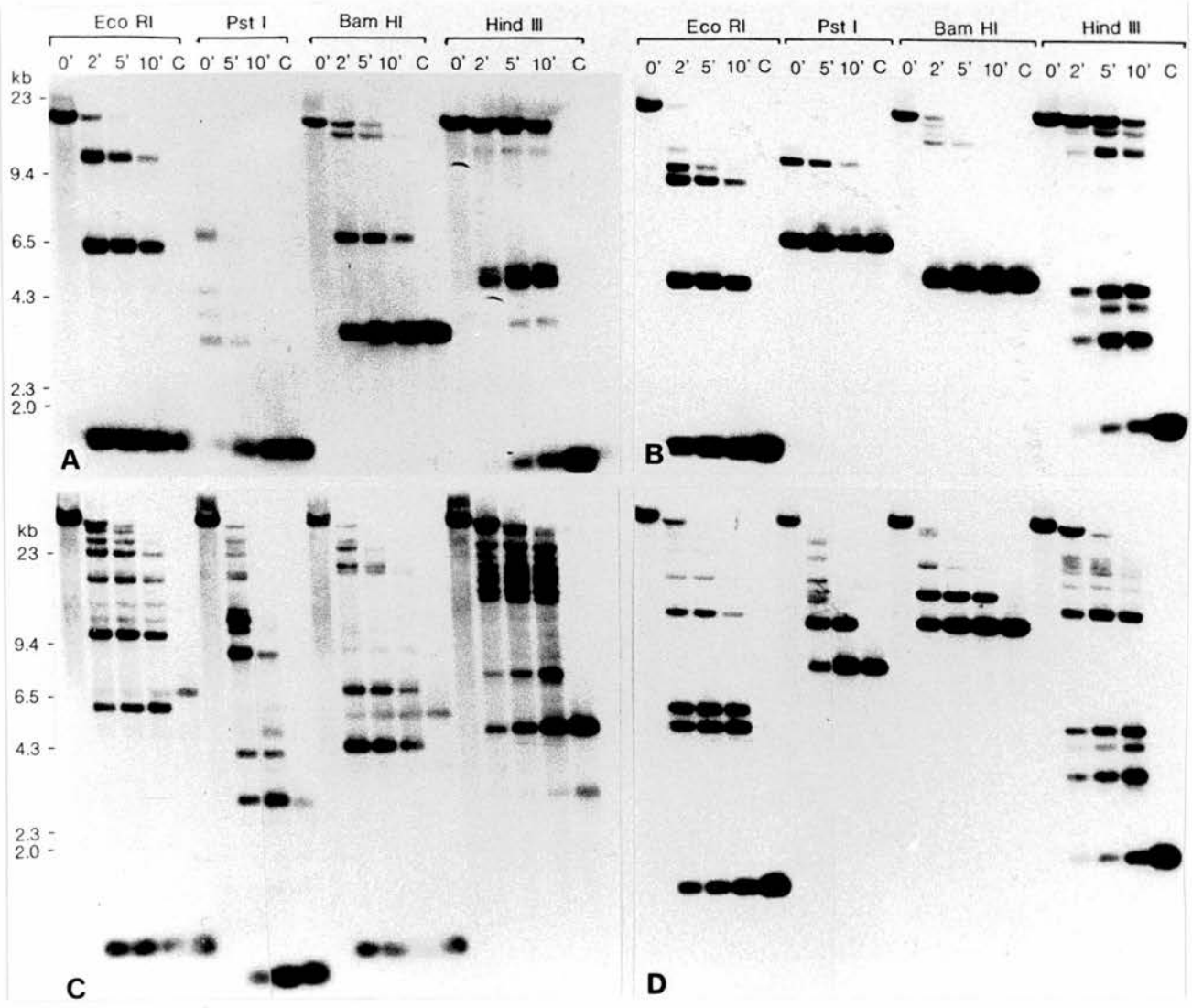


Figure 5.2 Southern blot analysis of pCB3 (A,B) and pCB4 (C,D) partial digests. The blots were sequentially hybridised with the BS (A,C) or the SA (B,D) end-specific probes.

Figure 5.3 Restriction maps of pCB3 and pCB4. Maps were produced using data from the end-specific probes hybridised to partial digests. The asterisk (*) denotes the *Pst* I site within the homeobox. The maps are aligned in order to demonstrate the overlap. Cosmids are linearised at the *Sal* I site in the vector. Vector sequences are indicated by the hatched box, to which the SA probe hybridises, and the closed box to which the BS probe hybridises.

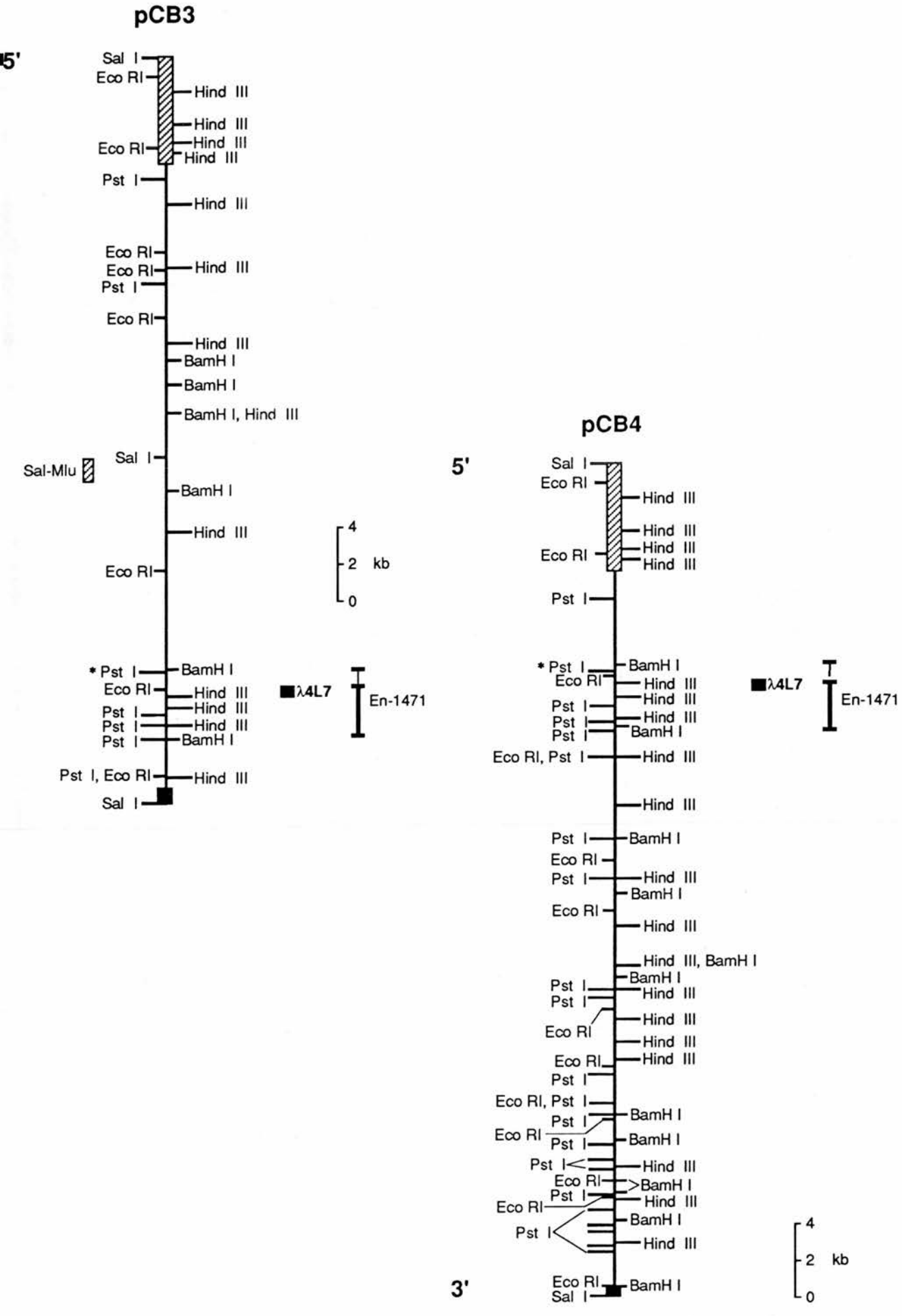
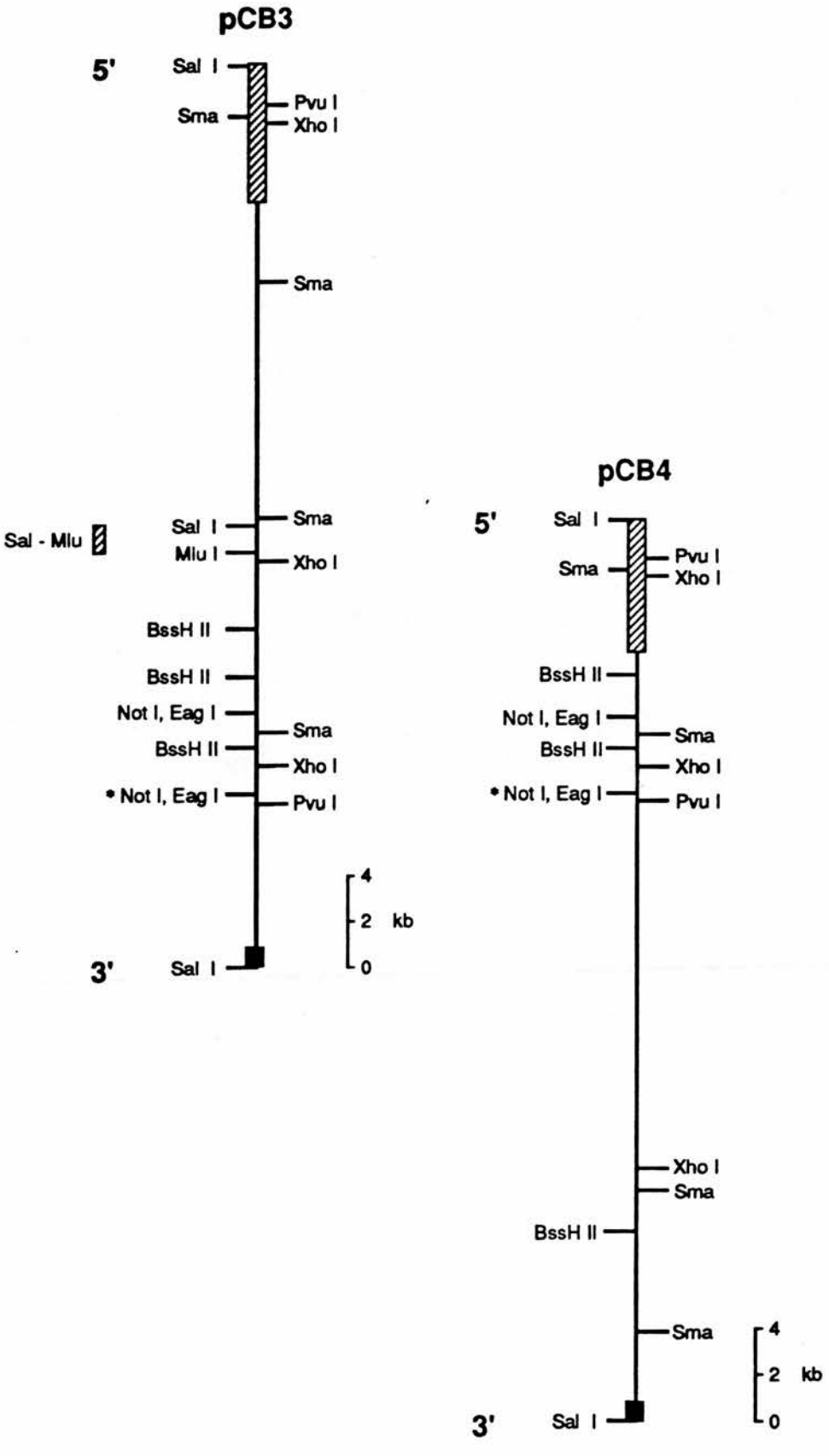

pCB3
5'
Sal I
Eco RI
Hind III
Hind III
Eco RI
Hind III
Hind III
Pst I
Hind III
Eco RI
Eco RI
Hind III
Pst I
Eco RI
Hind III
BamH I
BamH I
BamH I, Hind III
Sal I
Sal-Mlu
BamH I
Hind III
Eco RI
4
2 kb
0
* Pst I
BamH I
Eco RI
Hind III
λ4L7
Hind III
Pst I
Hind III
Pst I
Hind III
Pst I
BamH I
En-1471
Pst I, Eco RI
Hind III
Sal I
pCB4
5'
Sal I
Eco RI
Hind III
Hind III
Hind III
Eco RI
Hind III
Pst I
* Pst I
BamH I
Eco RI
Hind III
λ4L7
Hind III
Pst I
Hind III
Pst I
BamH I
En-1471
Eco RI, Pst I
Hind III
Hind III
Pst I
BamH I
Eco RI
Pst I
Hind III
BamH I
Eco RI
Hind III
Hind III, BamH I
BamH I
Pst I
Hind III
Pst I
Hind III
Eco RI
Hind III
Eco RI
Hind III
Pst I
Eco RI, Pst I
Pst I
BamH I
Eco RI
Pst I
BamH I
Pst I
Hind III
Eco RI
BamH I
Pst I
Hind III
Eco RI
Pst I
BamH I
Hind III
4
2 kb
0
3'
Eco RI
Sal I
BamH I

Figure 5.4 Restriction maps of pCB3 and pCB4. Maps were produced using data from the end-specific probes hybridised to complete and partial digests with rare cutting enzymes. The asterisk (*) denotes the *Not* I site within the homeobox. The maps are aligned in order to demonstrate the overlap. Cosmids are linearised at the *Sal* I site in the vector. Vector sequences are indicated by the hatched box, to which the SA probe hybridises, and the closed box to which the BS probe hybridises.


pCB3
5'
Sal I
Sma
Pvu I
Xho I
Sma
Sal - Mlu
Sal I
Mlu I
Sma
Xho I
BssH II
BssH II
Not I, Eag I
Sma
BssH II
Xho I
* Not I, Eag I
Pvu I
4
2
kb
0
3'
Sal I
pCB4
5'
Sal I
Sma
Pvu I
Xho I
BssH II
Not I, Eag I
Sma
BssH II
Xho I
* Not I, Eag I
Pvu I
Xho I
Sma
BssH II
Sma
4
2
kb
0
3'
Sal I

5.2.4 *Subcloning the Sal I - Mlu I Fragment for use as a 5' mapping probe*

As discussed in the introduction, in order to extend the long range map bidirectionally it is necessary to (1) know the location of rare cutting enzymes sites within the *En-1* gene and (2) to have probes which hybridise to either side of these sites. The mapping of the cosmids, as described above, detected the presence of several rare cutting enzymes within the *En-1* gene. From the map (see Fig. 5.4) it can be seen that λ4L7, which is the *En-1* specific probe from the *En-1* cDNA, hybridises 3' to these rare cutting enzyme sites. From the map of pCB3 (Fig. 5.4) it can be seen that there is a 1.2kb *Sal* I-*Mlu* I fragment located 5' to the rare cutting enzyme sites. This fragment was cloned into pBLRC (see Fig. 6.4a, Chapter six). The isolated *Sal* I - *Mlu* I fragment was then hybridised to Southern blots of both cosmids, to ensure that it was indeed pCB3 specific and did not contain any repeat elements.

5.2.5 *Subcloning the Bam HI Fragment for use as a 3' mapping probe*

Initial PFGE studies were carried out using λ4L7 as a probe. This probe however repeatedly failed to be informative. Therefore the *Bam* HI fragment which included the homeobox and extended 3' (denoted as En-1471 in Fig. 5.3) was subcloned into the vector pTZ18U. Positive colonies were detected via hybridisation to an oligonucleotide (oligo 471) which hybridises to sequences 3' of the homeobox (see Fig. 5.5). The oligonucleotide 471 was synthesised using sequence information from Joyner & Martin (1987). The insert was released from the plasmid via digestion with *Eco* RI and *Bam* HI releasing a 700bp fragment, which spans the homeobox, and a 2.6kb fragment which is *En-1* specific (see Fig.5.3). The 2.6kb insert which was called pEn-1471 was subsequently used for both PFGE mapping and analysis of recombinant offspring in the linkage analysis (see Chapter four, Fig 4.1)

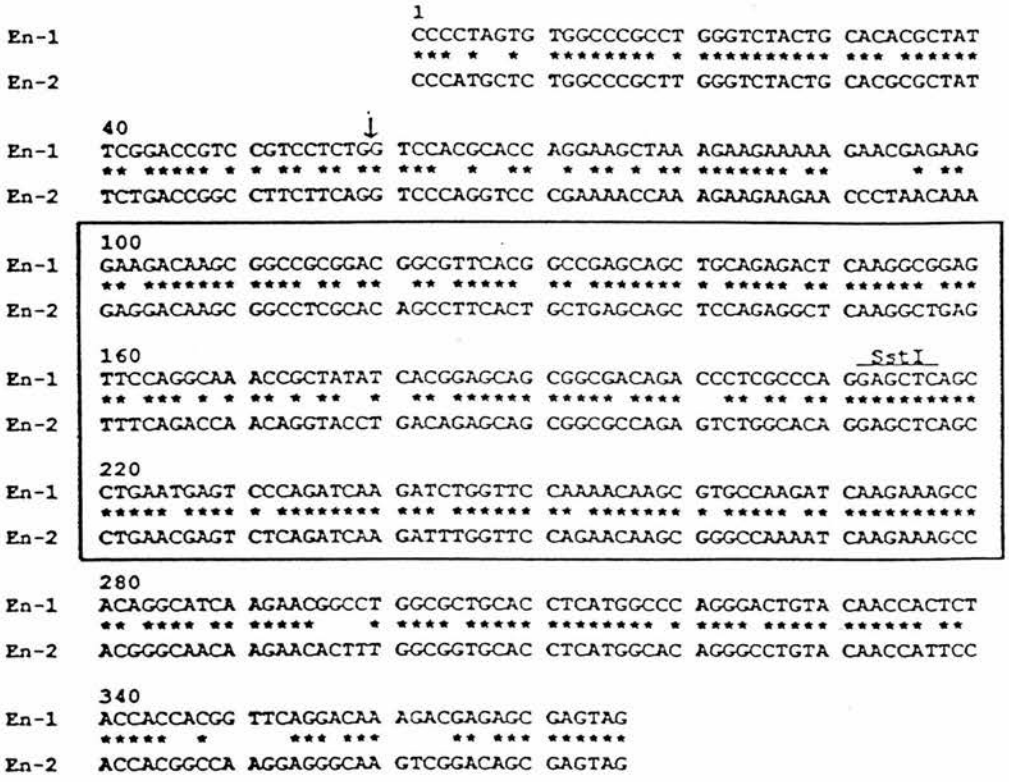


Figure 5.5 Sequence analysis of the conserved region in *En-1* and *En-2*. For both genes the arrow indicates the boundaries of an intron, and the box delineates the homeobox. Asterisks denote nucleotides that are identical in *En-1* and *En-2*. The last three nucleotides encode a termination signal for the translation of the open reading frame. The oligonucleotide 471 hybridises to the last 17 nucleotides of the *En-1* sequence.. Figure reproduced from Joyner & Martin (1987).

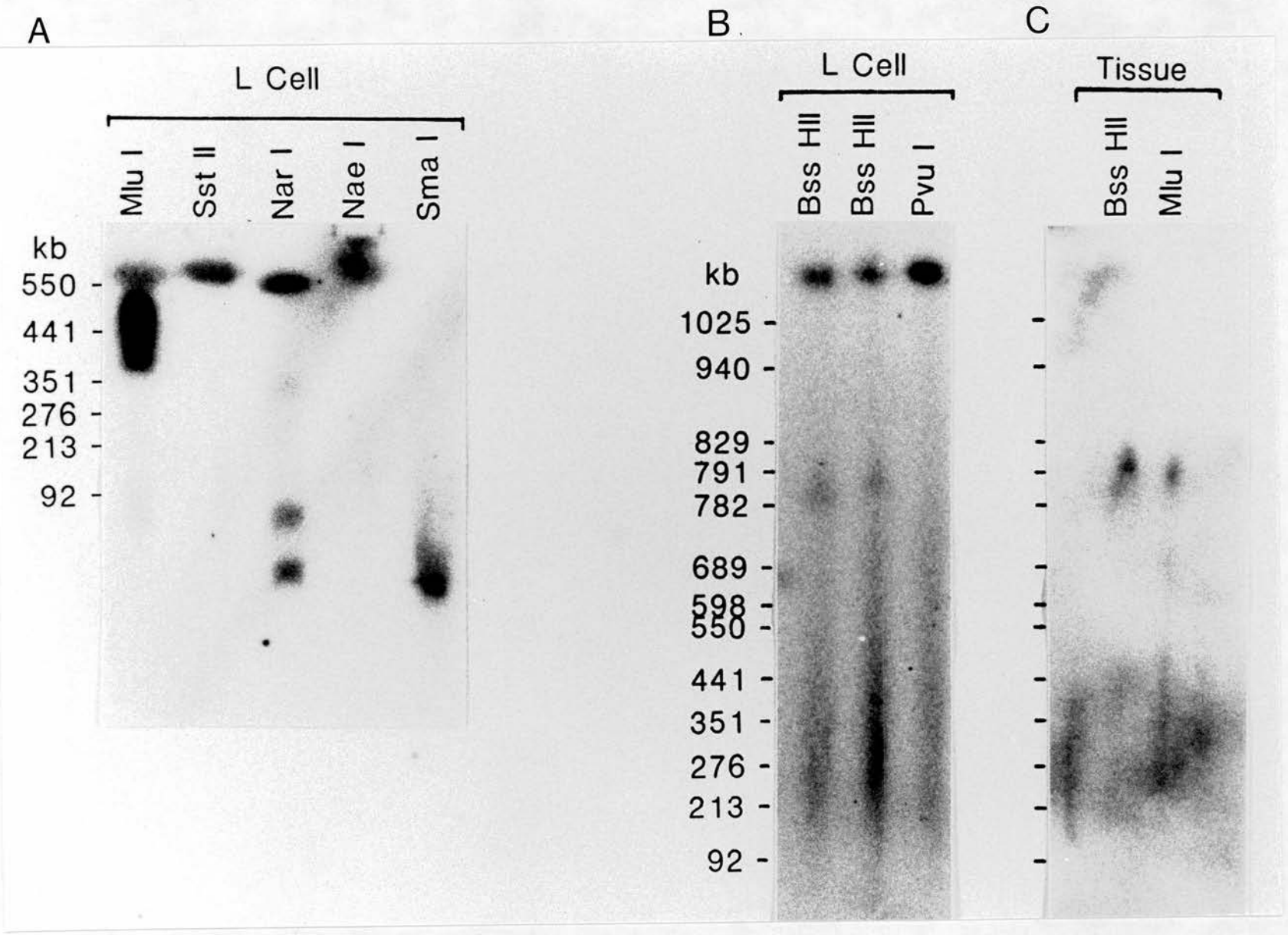
5.2.6 *Pulsed Field Gel Electrophoresis*

Initial PFGE studies were performed with DNA from the murine embryonic L cell line, using the pλ4L7 probe. This probe hybridised to a 800kb *Bss* HII fragment, a 200kb *Not* I fragment and, rather than hybridising to a discrete fragment with the enzyme *Mlu* I, pλ4L7 produces a strong signal which spans 150kb. This smear was mapped to approximately 400-550kb 3' to *En-1*. The presence of these bands were later confirmed in DBA and C57BL/6 mouse strain DNA. However, rather then obtaining the 150kb smear with the enzyme *Mlu* I, pλ4L7 hybridises to a 800kb band which co-migrates with the 800kb *Bss* HII band (see Fig. 5.6).

One of the major limitations at the start of this mapping work was the fact that only λ4L7 was available as a mapping probe. Initially a 3kb region of DNA upstream to the most 5' *Not* I site of *En-1* site was subcloned from the genomic λ clones (data not shown). However this probe was found to contain repeat sequences and was therefore unsuitable for mapping studies. This problem was overcome by the isolation and subsequent subcloning of the *Sal* I - *Mlu* I fragment (see above) which was used as a 5' mapping probe (see Fig. 5.7). This probe hybridised to a 600kb *Not* I fragment, a 500kb *Sst* II and a 350kb *Eag* I fragment. This probe was also found to hybridise to fragments generated with several other enzymes.

The resulting long range PFGE map around *En-1* is given in Fig. 5.8.

Figure 5.6 Southern blot analysis of L cell and tissue DNA using the pλ4L7 probe. *(A)* pλ4L7 hybridises to a *Mlu* I smear which spans 150kb and (*B*) a 790kb *Bss* HII fragment in L cell DNA. In tissue DNA, (*C*), pλ4L7 hybridises to a 790kb *Bss* HII and a 790kb *Mlu* I fragment.


A
L Cell
Mlu I
Sst II
Nar I
Nae I
Sma I
kb
550
441
351
276
213
92
B
L Cell
Bss HII
Bss HII
Pvu I
kb
1025
940
829
791
782
689
598
550
441
351
276
213
92
C
Tissue
Bss HII
Mlu I

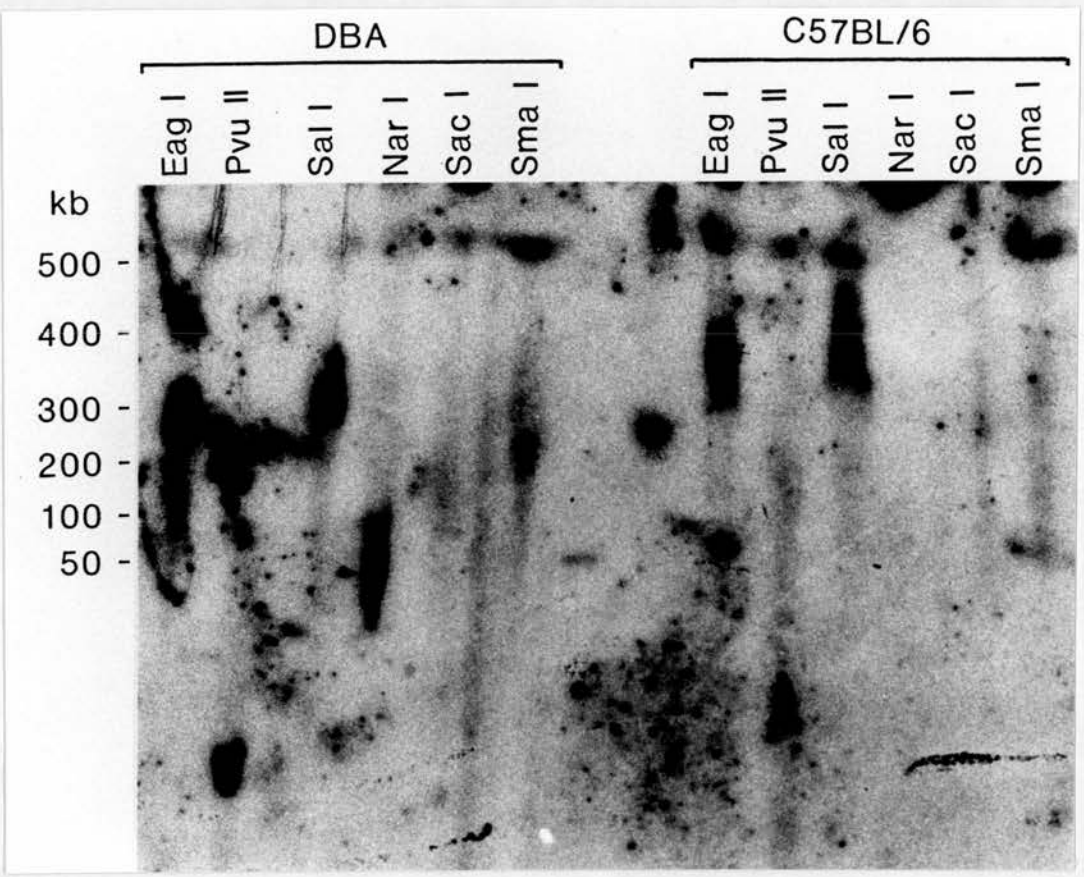


Figure 5.7 Southern blot analysis of DBA and C57BL/6 DNA hybridised to the Sal-Mlu probe. This probe detected the presence of *Sal* I and *Eag* I (~350kb) and *Pvu* II (<50kb) sites in both DNAs. The Sal-Mlu probe also hybridises to a 50kb Nar I band in DBA DNA.

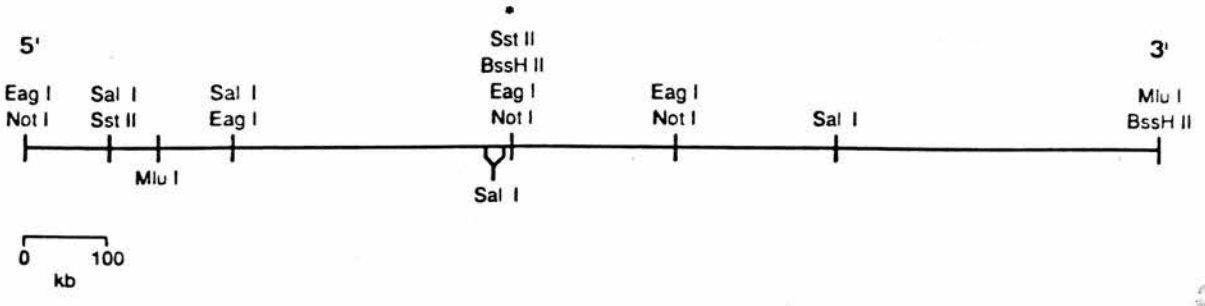


Figure 5.8 Long range restriction map of DBA and C57BL/6 DNA. The map was produced using data generated from pulsed field gels hybridised to the pλ4L7, pEn-1471 and Sal-Mlu probes. The asterisk denotes the position of those sites clustered around the 5' end of *En-1*.

5.2.7 *Long Range Partial Mapping*

The mice recovered from the linkage analysis which are recombinant between *En-1* and *Dh*, do not in themselves yield precise information with regard to the physical position of the *Dh* gene relative to *En-1*. One method of localising the *Dh* gene and orientating *En-1* relative to the chromosome would be to locate the recombination breakpoints in these mice. It was hoped to achieve this via mapping hybridisation patterns of partial restriction digests. This type of analysis takes advantage of the data provided in both the short and long range maps and the availibility of the two probes (Sal-Mlu and pEn-1471; see sections 5.2.4 and 5.2.5) which flank *En-1*.

The most distant mapped sites from *En-1* are the 600kb 5' *Not* I site and the 800kb 3' *Bss* HII fragments (see fig. 5.8). The Sal-Mlu probe hybridises to the *Not* I fragment and the En-1471 probe hybridises to the *Bss* HII fragment. If one digests DNA to completion with Not I and then performs a second partial digest with a more frequently cutting enzyme the Sal-Mlu probe will hybridise to the 600kb Not I and a "ladder" of partial restriction products which extend from the En-1 gene towards the 3' Not I site. Similarly if one digests the DNA to completion with Bss HII and hybridises with the En-1471 probe, a ladder of fragments which extend from the En-1 gene towards the 800kb Bss HII site will be detected. This type of analysis will therefore produce a high resolution map for the second more frequently cutting enzyme. To orientate the *En-1* gene the second enzyme will be one which, in FZT and C57BL/6 DNA generates a polymorphic ladder of fragments. In recombinant animals one would expect to see a pattern of partial restriction fragments representing C57BL/6 DNA followed by a transition into the ladder of partial fragments representative of FZT DNA or vice versa. The site at which the transition occurs is the region of the recombination breakpoint.

Enzymes which contain the CpG dinucleotide in their recognition sequence were chosen to look for a polymorphism. These enzymes vary with regard to the size of the average restriction fragment which they generate in mammalian DNA, and range from an average fragment size of 100kb to 1000kb. Enzymes such as *Sal* I, *Cla* I and *Xho* I, which generate fragments of an average size of 100kb, were chosen to look for a polymorphism. Therefore, although these enzymes are rare cutters, they cut frequently enough to produce a more precise location of the recombination breakpoint than an enzyme which cuts on average every 500kb. This analysis will have the added benefit of increasing the resolution of the long range map.

As the orientation of En-1 relative to the chromosome was unknown it was necessary to extend the partial mapping in both directions. As both 5' (Sal-Mlu) and 3'(pEn-1471) probes were now available, it was theoretically possible to orientate partial maps by cutting the DNA to completion with *Not* I, yielding a 600kb fragment 5' and a 200kb fragment 3' to *En-1*. Partial digests of DNA blocks were performed by altering the salt concentration of the restriction buffer rather than altering the time of digestion or using dilute enzyme due to the fact that once enzyme penetrated the agarose block it would be difficult to accurately stop the digestion reaction. Although a partial pattern was observed with the Sal-Mlu probe after complete digestion with *Not* I, followed by partial *Sal* I digestion, no polymorphism between DBA and C57BL/6 DNA was found (see Fig. 5.9). Attempts to find a polymorphic hybridisation pattern using either probe and a variety of enzymes either failed to determine any polymorphism or resulted in non-informative gels.

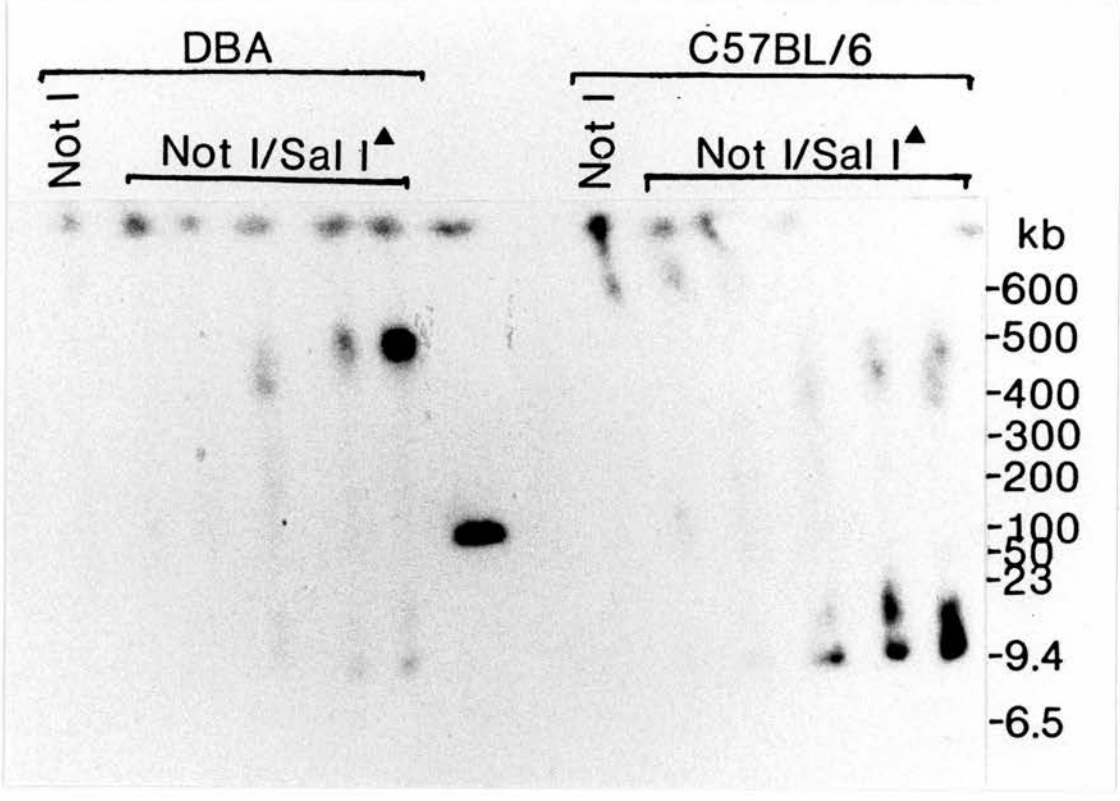


Figure 5.9 Southern blot analysis of long range partial mapping of DBA and C57BL/6 DNA. The DNA was first digested to completion with *Not* I before being partially digested with *Sal* I (denoted as Not I/Sal I) and probed with the Sal-Mlu probe. Partial digests were performed by increasing the salt concentration of the restriction buffer. The signal seen in lane 7 is yeast marker; residual plasmid sequences attached to the probe cross hybridise with the 92kb yeast chromosome.

5.2.8 *Long range mapping of Dh DNA.*

The long range mapping described above focused on the localisation of neighbouring genes and establishing the recombination breakpoint of the four animals recovered from the linkage analysis. Another important approach to the localisation of the *Dh* locus is to determine whether one can detect any chromosomal rearrangements in this region of the genome. If the *Dh* mutation is the result of a sizeable chromosomal rearrangement such as deletion or insertion then this could be detected via PFGE. PFGE on DNA from *Dh* homozygotes was performed. This analysis proved inconclusive.

5.3 Discussion

A long range map around *En-1*, which encompasses 1400kb, was established using enzymes which contain the dinucleotide CpG in their recognition sequence and cut rarely in genomic DNA. These enzymes are 'rare-cutters' for several reasons including the under representation of this dinucleotide in the genome (Razin & Riggs, 1980) and the inhibition of cleavage due to methylation of the cytosine residue in the enzyme recognition site (Brown & Bird, 1986). One other major property of the CpG dinucleotide is the clustering of non-methylated CpGs in G+C rich islands (Bird, 1986). This abundance of CpG dinucleotides is partially due to a high G+C content (65% compared to 40% in bulk DNA) but is mainly due to the lack of CpG methylation (Bird *et al.*, 1985; Tykocinski & Max, 1984).

Perhaps the most striking feature of these CpG islands is their association with genes. It is now quite clear that many CpG islands are associated with the 5' end of genes, particularly the region where transcription starts. One detailed study (Gardiner-Garden & Frommer, 1987) showed that all sequenced house-keeping genes transcribed by RNA polymerase II have 5' islands, although this was determined only by the number of CpG residues, lack of methylation was not considered. Several tissue specific genes have also been shown to have CpG islands eg Thy-1 gene of mice (Spanopoulou *et al.*, 1986) and the α-globin genes of man (Bird *et al.*, 1987).

CpG islands can be observed in a PFGE map as a clustering of rare-cutter sites (Brown & Bird, 1986). From the cosmid map (Fig 5.4; Fig 5.8) it can be seen that there is a clustering of *Not* I, *Eag* I, *Sst* II and *Bss* HII around the 5' region of *En-1* all of which have a high association with CpG islands. This would suggest, on the basis of the presence of CpG dinucleotide alone, that *En-1* has a 5' CpG island.

Initially PFGE was performed on DNA from the murine embryonic L cell line. Using the 3' probe λ4L7 three sites were mapped: *Not* I, *Bss* HII and *Mlu* I. Both *Not* I and *Bss* HII have a high association with CpG islands: 90% of all *Not* I sites and 75% of all *Bss* HII sites are found within islands (Bird, 1987). Although a lower percentage of *Bss* HII sites are found in islands, those which are inter island are usually methylated in genomic DNA (W. Bickmore,unpublished results). *Mlu* I, on the other hand, has a very low association with islands, as only 27% of all *Mlu* I sites are associated with islands (Bird, 1987; W Bickmore, unpublished results). The recognition sites for most of these rare cutting enzymes contain the CpG dinucleotide and the methylation of cytosine residues in the genome can therefore potentially prevent restriction (Smith *et al.*, 1987b). A variation in the methylation status of these sites in cell lines and tissue can therefore result in a difference in the size of restriction fragments and thus in the map. Partial methylation may result in incomplete digestion. An example of this phenomenon is seen with *Mlu* I. In L cell DNA digested with *Mlu* I pλ4L7 hybridises to a 150kb smear, which is probably the result of partial methylation in this region of the genome. However in DBA and C57BL/6 liver DNA pλ4L7 hybridises to a 800kb *Mlu* I fragment. This suggests that in adult liver tissue this region of the genome is completely methylated.

The obvious question with regard to the *Not* I and *Bss* HII sites mapped is whether these sites represent islands. Both of these enzymes are highly indicative of islands and the detection of the same size fragments in both L cells and tissue is particularly encouraging as it further suggests that these sites are within islands. This would suggest that there are two independent islands on the 3' side of *En-1*. However, Antequera *et al.* (1990) found that L cells in culture tend to methylate and therefore shut down transcription of non-essential genes. This suggests that these

islands might be associated with house-keeping genes rather than a development specific gene such as *Dh*.

On the 5' side of *En-1* there are three possible islands. The closest of these has an *Eag* I site, the next a *Sst* II site and the furthest has a *Not* I site. All of these enzyme sites have a high association with islands (75% of all *Eag* I and *Sst* II sites are found in islands; 90% of all *Not* I sites are found in islands; Bird, 1987). Although *Not* I sites are therefore highly indicative of an island, not all islands have a site for this enzyme (W. Bickmore, unpublished results) which might explain the absence of a *Not* I site next to the *Eag* I and *Sst* II sites described. The majority of islands have sites for *Eag* I and *Sst* II, and inter-island sites are usually methylated (W. Bickmore, unpublished results), an observation that would suggest that the sites mapped around *En-1* do indeed represent islands. Overall these results therefore suggest that there are three islands 5' to *En-1*.

Thus this mapping data suggests there are a possible five CpG islands surrounding *En-1*, spanning a region 600kb 5' of the gene and 800kb 3' to it (Fig 5.8). The obvious question with regard to these islands is whether any of them could be associated with the *Dh* gene. In the mouse, as discussed in chapter four, 1cM = 2Mb although there have been reports of this average showing a wide variation (Barsh & Epstein, 19??; L. Stubbs, pers. comm.). If one assumes that the region around *En-1* adheres to the 1cM = 2Mb average then *Dh* and *En-1* will be 1.4± 0.8 Mb apart which suggests that the islands are not associated with *Dh*. On the other hand there is always the possibility that in this region of the genome this relationship could be disturbed as in the agouti region where 1cM was found to be equivalent to as little as 100-200kb. If this were the case then any one of the islands could represent *Dh*, although the 3' islands are more likely to be associated with house-keeping genes (see above) rather than a development specific gene such as *Dh* . One however must be aware that the

physical mapping of CpG islands is not a definitve method of localising all genes as not every gene has an associated island (Gardiner-Garden & Frommer, 1987). However, it has been estimated that there are 30,000 CpG islands in the vertebrate genome (Bird *et al.*,1985) and as it has been anticipated that there are 20,000-50,000 genes in the vertebrate genome this would therefore suggest that a high proportion of all islands will be gene associated.

The four *Dh* - *En-1* recombinant animals recovered from the linkage analysis provide an extremely valuable resource for clarifying the position of the *Dh* locus. Mapping by hybridisation of partial restriction fragments showing a polymorphism between DBA and C57BL/6 DNA will facilitate the indentification of the recombination breakpoint in each mouse. Attempts to find a polymorphic hybridisation pattern between DBA and C57BL/6 DNA have so far been unsuccessful.This was in the main due to the majority of gels being uninformative rather than the detection of non-polymorphic hybridisation patterns.

A time consuming aspect of the mapping studies is that although it is known that *Dh* is separable by recombination from sequences at the 3' end of *En-1* (see Chapter four), the orientation of *En-1* relative to the chromosome is unknown. If the orientation of *En-1* could be determined then future mapping and cloning techniques need only be confined to one direction. Attempts were made to link, via PFGE, the molecular probes for *En-1* and *Emv-17*. This however proved unsuccessful. The isolation of flanking probes more distant from *En-1* will help in the orientation of *En-1* relative to the chromosome as will localisation of the recombination breakpoints. The next chapter describes an approach for the isolation of large cloned fragments around *En-1* which will have the benefit of yielding such flanking probes.

The long range mapping described in this chapter was performed using the pλ4L7, pEn-1471 and Sal-Mlu probes. As *Dh* has been shown to genetically map

between the *Emv-17* and *En-1* loci (see previous chapter) it is certainly worth establishing a long range map which extends from *Emv-17* towards *En-1*. This would first require the orientation of *Emv-17* relative to the chromosome and establishing a short range map around this locus. The isolation of probes between the *Emv-17* and *En-1* loci would greatly aid the mapping of this region. Indeed the isolation of probes more distant than those at present available is vital for the extension of the long range map, particularly in view of the fact that if this region of the genome adheres to the 1cM=2Mb average then *Dh* is approximately 1.4Mb proximal to *En-1*.

Although the application of PFGE is a very powerful method for the physical mapping of genes, one major disadvantage of this technique is that it does not provide any cloned DNA. The next chapter describes an approach for the isolation of large cloned fragments of DNA around *En-1* which will have the benefit of generating flanking DNA probes vital for the extension of the mapping studies described in this chapter.

Chapter Six

Construction of *En-1* Targeting Vector

6.1 Introduction

The previous two chapters dealt with the genetic and physical maps which have been built up around the *Dh - En-1* region. This chapter describes work which exploits both sets of data in order to aid the cloning of the *Dh* gene.

The combination of genetic maps, which are based on linkage data, and physical maps, which are based on cloned segments of DNA, is of great importance when studying any region of a chromosome. As is evident with the *Dh - En-1* region, genetic maps can resolve loci to within 1 cM of each other, although in physical terms this could represent as much as 4Mb (see previous chapter). Until quite recently this level of resolution was well outwith the the range of molecular cloning techniques most of which involved the use of λ replacement or cosmid vectors with a limited insert size of 24kb to 50kb respectively. Although it is possible to 'walk' along the chromosome by using the end of one clone to isolate adjacent overlapping clones (Bender *et al.*, 1983), each step is usually of a length less than the average insert size. To walk 1Mb would therefore require at least 20 steps, and in real terms a walk of this magnitude would doubtless require a minimum of 40 steps. This technique is therefore very labour intensive and even then the distance spanned may not exceed 200kb due to the added complication that some regions of mammalian DNA are unclonable in *E. coli* due to repeated DNA sequences and secondary structure (Wyman *et al.*, 1985). The use of jumping and linking libraries (reviewed by Poustka & Lehrach, 1986) can help "jump" over these unclonable regions but such libraries are of limited use as they only isolate DNA from the "take-off" and "landing points" of the jump. Likewise although PFGE can yield a tremendous amount of information in terms of a long range physical map the technique in itself does not yield cloned DNA.

The advent of YAC (yeast artificial chromosomes) cloning technology (Burke *et al.*, 1987) promises to help bridge the gap between genetic and physical maps. YAC

cloning techniques involve the ligation of large fragments of genomic DNA between two vector arms each of which ends in a telomere and is functional in S. cerevisiae. These vectors also contain a functional autonomous replicating sequence (ARS 1), a functional centromere (CEN 4), and two yeast selectable markers, one for each vector 'arm' (see Fig 6.1). There have been reports of insert sizes of 600kbp (Coulson *et al.*, 1988) and even up to 2Mb (L. Stubbs, pers. comm.). Besides this size advantage over lambda/cosmid vector inserts the use of YAC vectors may enable the cloning and maintenance of those sequences which are poorly represented in λ and cosmid libraries due to the previously mentioned unclonable DNA sequences (Coulson *et al.*, 1988).

One major drawback to the YAC vector system is that the actual construction of the library is extremely labour intensive. Each yeast colony needs to be hand picked and grown, to achieve a library of one genome equivalent, with an average insert size of 200kb, fifteen thousand clones would need to be picked. In order to enhance the probability that all regions of the genome will be cloned it is necessary to pick more than one genome equivalent otherwise contiguous (contig) mapping via overlapping clones and consequently the isolation of prospective genes may be hindered. One other major drawback involves the actual screening of the library. One possible method of screening involves the polymerase chain reaction (PCR). PCR is a very sensitive *in vitro* DNA amplification method (Saiki *et al.*,1988). If one has DNA sequence information for the region of interest then it would be possible to synthesise oligonucleotides primers and amplify the target DNA. The huge task of screening the library could be reduced by initally screening pools of YAC clones. However contamination with DNA can be a serious problem with PCR and carry-over of even minute quantities of reaction product can lead to serious contamination thereby resulting in false positives. Alternatively YAC libraries can be plated out onto Nitrocellulose / Hybond membranes in a grid like manner and probed radioactively.

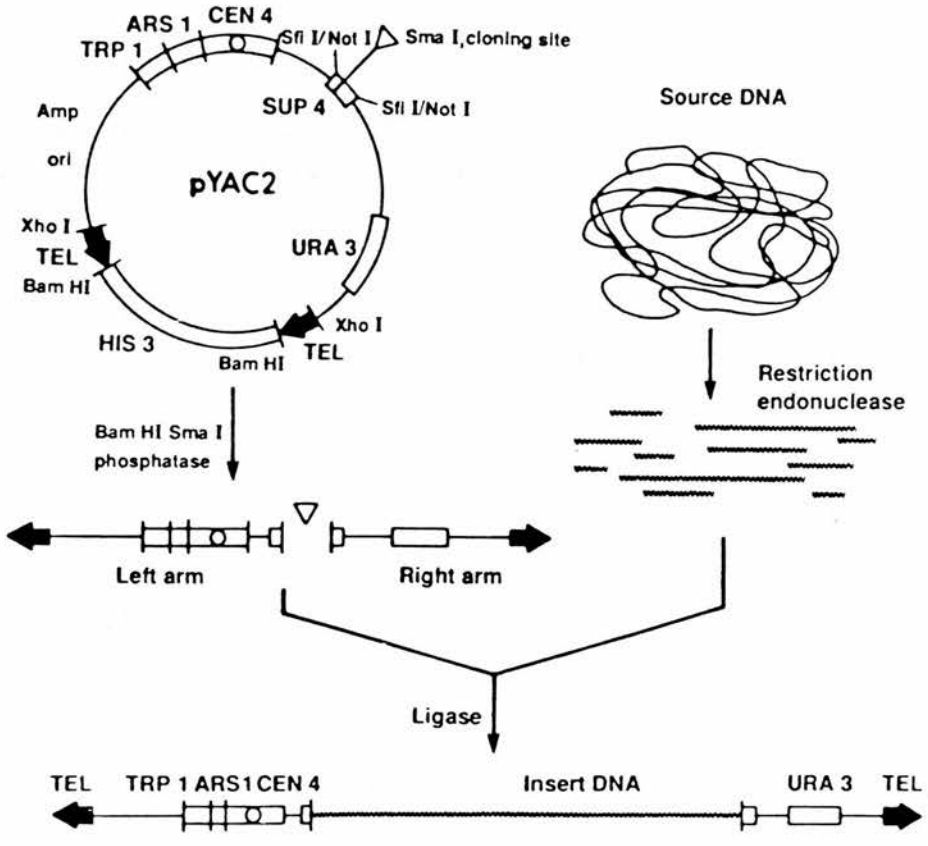


Figure 6.1 Yeast artificial chromosome (YAC) cloning system. pBR322 derived sequences are indicated as a thin line. SUP4, TRP1, HIS3 and URA4 are yeast genes. SUP4 is an ochre suppressing allele of a tyrosine transfer RNA gene that is interupted when foreign DNA is cloned into the vector. TRP1 and URA4 allow for the selection of molecules that have acquired both chromosome arms from the vector. HIS3 is discarded during the cloning process. ARS1 is an autonomously replicating sequence, CEN4 provides centromere function. Within the context of this project the source of DNA for the construction of the library will be DNA from *En-1* targeted RAG cells. (Reproduced from Burke *et al.*, 1987).

This would require griding every colony and screening every filter. The major problem with this method is that the filters have a limited life span and therefore repeated screening of the library would require fresh filters.

The work presented in this chapter describes the construction of a vector which aims to allow the construction of small YAC libraries which are enriched for the region of the genome containing *Dh* and *En-1*. The vector has been modeled on the gene targeting vectors described by Thomas & Capecchi (1987) and Mansour *et al.* (1988). Gene targeting is the term given to the homologous recombination between DNA sequences in the chomosome and newly introduced DNA sequences. Thus *in vitro* alterations in cloned DNA sequences could be targeted into the genome. In this case the *En-1* gene will be interrupted with two selectable markers and introduced into mouse cells (see later).

The vector described in this chapter contains approximately 11 kb of *En -1* (see Fig.6.5), the target gene, within which two selectable markers have been cloned. The herpes simplex virus gene thymidine kinase (HSV-*tk*) has been cloned adjacent to *En-1*. Of the two markers cloned within the *En-1* region one is the neomycin resistance gene (*neo* ^{r}) which is the mammalian selectable marker for G418 resistance (G418^{r}). The other marker is LEU2 which is a yeast selectable marker. Obviously the requirement of a selectable marker is absolute and the cloning of the *neo*^{r} and *tk* genes as described above are based on the method of positive negative selection (PNS) which greatly enhances the targeting event (Mansour *et al.*, 1988; see Fig 6.2). PNS uses a positive selection, in this case neomycin resistance, for cells that have incorporated the vector anywhere in their genome coupled with a negative selection, in this case HSV-*tk* , against those cells that have random integrations of the vector within their genome. Only those cells which contain the neomycin resistance gene can grow in the presence of the drug G418. Cells which contain the HSV-*tk* gene will be

killed by gancyclovir. Correctly targeted cells will be *neo* $^{r+}$, HSV-*tk* $^{-}$ and therefore resistant to both G418 and gancyclovir. The purpose of LEU 2, the yeast selectable marker, is described below.

The overall stategy for the work described in this chapter is as follows. The first stage in the construction of small YAC libraries enriched for the *En-1* and *Dh* region will be the targeting of the finished construct into the mouse RAG cell line. The stepwise construction of the targeting vector is represented schematically in Fig.6.3. The RAG cell line is derived from a murine mammary tumour and is susceptible to selection via neomycin resistance (Sinead Jones, pers. comm.). Once targeted cells have been identified (*neo*$^{r+}$, *tk*$^{-}$ and LEU2^{+}), DNA from these cells will be used as a source of DNA to make a YAC library. Transformation of the resultant clones onto a Leu2^{-} background strain of yeast will facilitate the preferential selection of those clones which contain *En-1* and flanking DNA.

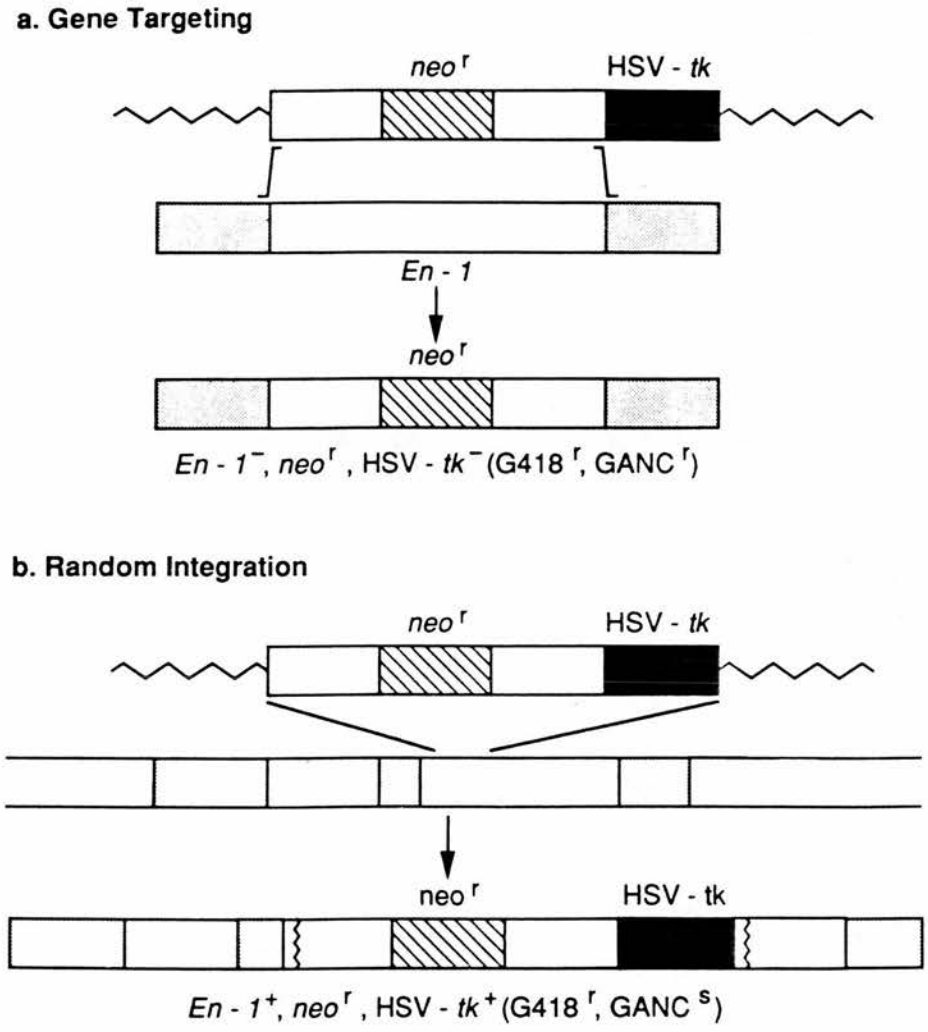


Figure 6.2 Positive negative selection (PNS) procedure to enrich for cells containing a targeted disruption of *En-1*. (*a*) *En-1* replacement vector with the *neo*^{r} gene cloned within, and the HSV-*tk* gene cloned adjacent to, the *En-1* gene. Homologous recombination between vector and genomic *En-1* will result in the disruption of one copy of *En-1* and loss of the HSV-*tk* sequence. Targeted cells will be *En-1*$^-$, *neo*^{r}, HSV-*tk*$^-$ and therefore resistant to both GANC and G418. (*b*) Non-homologous integration of DNA into the genome occurs via the ends of linear DNA (Folger *et al.*, 1982; Roth *et al.*, 1985). In this instance the HSV-*tk* will remain linked to *En-1* / *neo*^{r}. Cells carrying a random integration will therefore be *En-1*$^+$, *neo*^{r}, HSV-*tk*$^+$ and therefore resistant to G418 but sensitive to GANC. (Adapted from Mansour *et al.*, 1988).

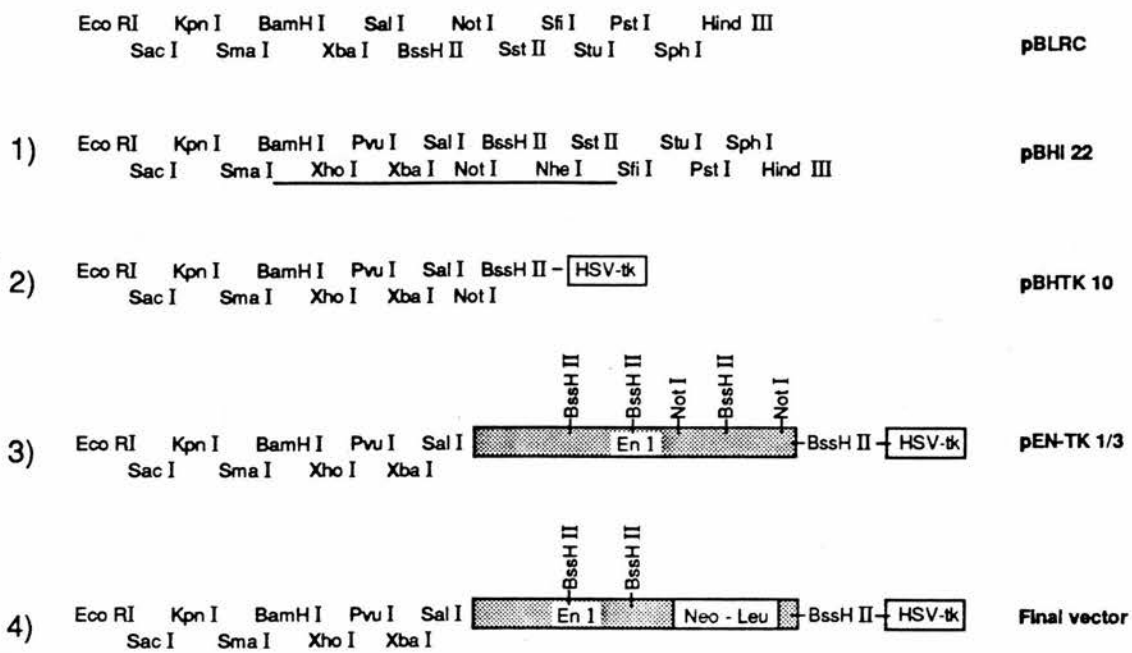


Figure 6.3 Schematic representation of the sequential cloning steps in the construction of the *En-1* targeting vector. (1) pBHI 22 was produced by replacing the *Bam* HI-*Sst* II fragment of pBLRC with the underlined *Bam* HI-*Sst* II fragment in pBHI 22. (2) HSV-*tk* was cloned between the *Nhe* I-*Hind* III sites to produce pBHTK 10. (3) *En-1* DNA with blunt end termini was cloned into the *Not* I site of pBHTK 10, which had also been treated to give blunt end termini, to generate pEn-TK 1/3. (4) Final vector, the Neo-Leu cassette will be cloned between the *Not* I sites within *En-1*. This step will be performed so that the *Not* I restriction sites will not be regenerated. The restriction sites shown above are only representative of the polylinker region of pBHI 22.

6.2 Results

The first step in the construction of this targeting vector was to find a plasmid which contained suitable sites for the various cloning steps. As no commercially available plasmids were suitable it was necessary to modify a pre-existing plasmid to fit the particular need of this construction.

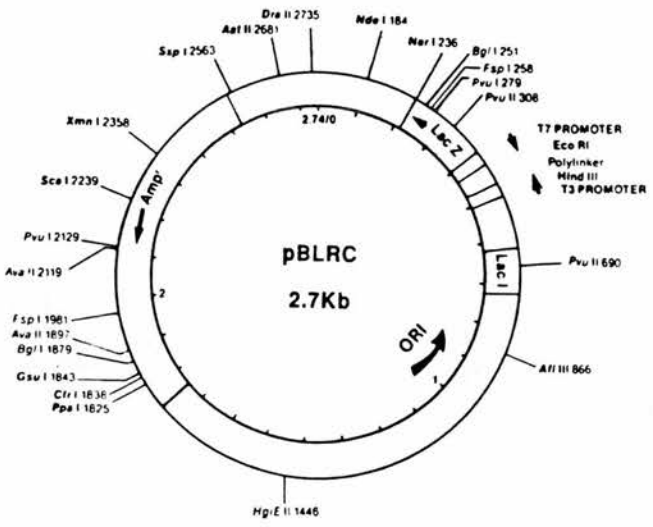
6.2.1 *Modification of the pBLRC plasmid*

pBLRC is a Bluescribe based plasmid into which oligos had previously been cloned, in-frame, into the polylinker region by Dr. R Hill see Fig 6.4a.

Two complementary oligonucleotides which had more convenient restriction enzyme sites were synthesised, annealed together and then cloned between the *Bam* HI and *Sst* II sites in the polylinker of pBLRC. These oligonucleotides were cloned inframe to retain β-galactosidase activity such that they replaced some original sites. The new plasmid which was called pBHI 22 is given in Fig 6.4b. It differs from pBLRC only in the polylinker region.

The *Pvu* I site in the oligonucleotide polylinker of pBHI22 cannot be used as a cloning site because there are two other *Pvu* I sites in pBHI22 (Fig.6.4b) and digestion with *Pvu* I would result in the linearisation of the plasmid. However *Pvu* I will allow linearisation of the final vector. It has been reported that targeting efficiences are greatly increased when linear rather than supercoiled DNA is used (Thomas *et al.*, 1986).

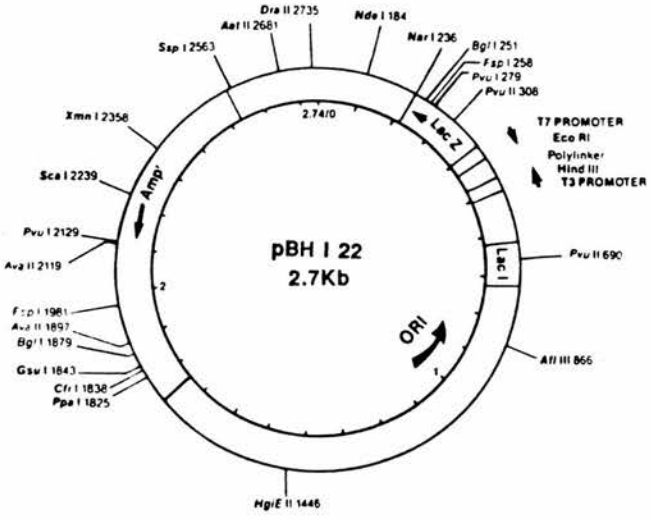
a



pBLRC polylinker

Eco RI Kpn I Xma I BamH I Sal I Not I Sfi I Pst I Hind III
Sac I Sma I Xba I Acc I BssH II Eag I Sst II Stu I Sph I
Hinc II

b



pBHI 22 polylinker

Eco RI Kpn I BamH I Pvu I Sal I BssH II Sst II Stu I Sph I
Sac I Sma I Xho I Xba I Not I Nhe I Sfi I Pst I Hind III

Figure 6.4 (*a*) pBLRC is the parental plasmid of the *En-1* targeting vector. (*b*) The pBHI 22 plasmid differs from pBLRC only in the polylinker region. The different restriction sites are underlined.

6.2.2 *Cloning of HSV-Tk Gene into pBHI22 (Production of pBHTk-10)*

The HSV-*tk* gene had previously been cloned into pUC by Dr. R Allshire such that its 5' end was orientated toward the *Eco* RI site. This plasmid was digested with the *Xba* I and *Hind* III and electrophoresed on low melting point (LMP) agarose. The gel slice containing the *tk* insert was cut out and treated with Gene-Clean as described in materials and methods.The *tk* insert was then cloned into the *Nhe* I / *Hind* III sites of pBHI 22 and then transformed into competent JM83. Positive colonies were streaked in triplicate onto nitrocellulose filters, one filter was kept as a master copy and the remaining two filters were processed using the Grunstein & Hogness method (see Materials & Methods, section 3.6) before being screened with oligo 383 which hybridises to HSV-*tk* sequences. One resultant clone, called pBHTk-10, was used for further manipulations. The *tk* gene was cloned in this orientation so that in the final linear vector the 5' end will be internal and therefore the upstream promoter will be protected from exonuclease activity. The *tk* gene represents a region of non-homology between the endogenous and exogenous target DNA and therefore, in correctly targeted cells, will not recombine homologously. Targeted cells will therefore be *neo* $^{r+}$ and HSV-*tk* $^{-}$ (G418^{r}, GANC^{r}).

6.2.3 *Cloning of En-1 Region of Homology into pBHTk-10 (Production of pEN-TK1 and pEN-TK3)*

The *Pvu I-Mlu I* fragment of the cosmid pCB3 was used as the region of *En-1* homology (Fig 6.5). This fragment is 11kb long and therefore represents a good length of homology with the endogenous region of DNA. Gene targeting frequencies are thought to be very sensitive to the extent of homology between the exogenous and

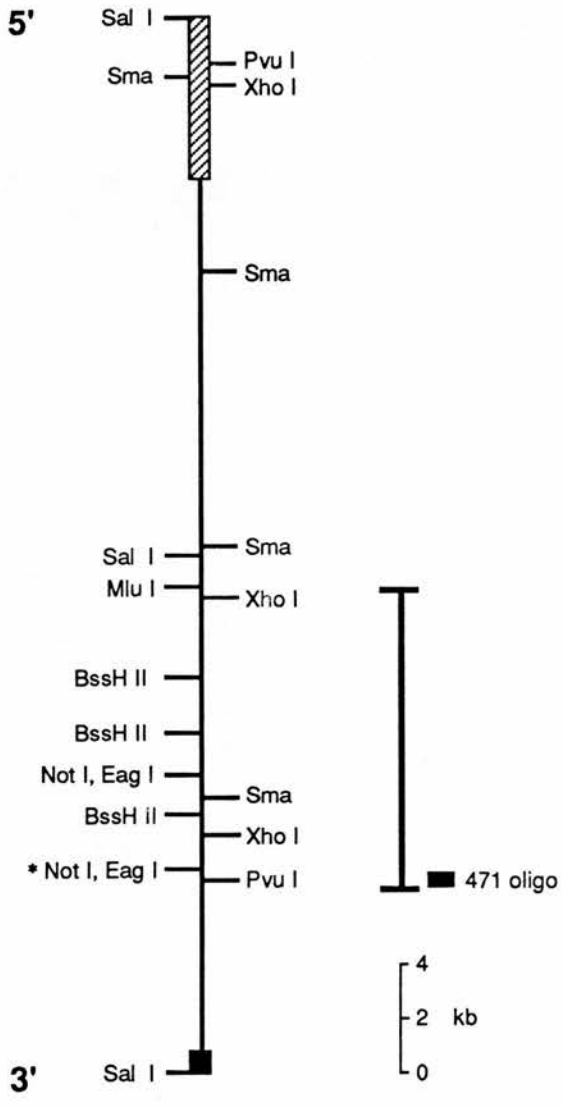


Figure 6.5 Restriction map of pCB3. The *Pvu* I-*Mlu* I fragment used in the construction of the targeting vector is indicated by the heavy bar. The region to which the 471 oligonucleotide hybridises is indicated. The *Not* I site situated within the homeobox is denoted by an asterisk.

endogenous sequence (Thomas & Capecchi, 1987). This length of DNA should enhance targeting frequencies without compromising the cloning ability of the plasmid

The cosmid pCB3 was digested to completion with *Pvu* I and *Mlu* I. The resultant cohesive ends were then treated with T4 DNA polymerase in order to generate blunt end termini. Aliquots of this were then 'shot-gun' cloned (i.e. all fragments produced during digestion were cloned) into pBHTk-10 (see above) which had been digested with *Not* I and then treated with T4 DNA polymerase. The ligation mix was transformed into competent JM83 and plated out onto a 20x20cm^2 Hybond membrane. Two replica filters were made from this master plate and all three plates were allowed to recover for 3-4 hours at 37^{0}C.The two replica filters were processed using the Grunstein & Hogness method (see Materials & Methods, section 3.6).One filter was screened with an oligonucleotide which hybridised to the *tk* gene (oligo 383) the other filter was screened with an *En -1* specific oligonucleotide (oligo 471). Two colonies were found to be positive for both probes and these were isolated, grown up and the plasmids recovered called pEN-TK1 and pEN-TK3. Restriction digestion with *Not* I confirmed that both plasmids contained the *En-1* region as both released the expected 3.5kb fragment (see Fig 6.6a). However as the cosmid fragment was subcloned into blunt-ended sites the orientation within each plasmid was unknown.The orientation of each insert was determined via restriction analysis and hybridisation. A unique *Sal I* site is present in the vector (see Fig 6.4b), pEN-TK1 and pEN-TK3 were digested in duplicate with either *Sal* I ,*Not* I or *Sal* I & *Not* I (neither of these enzymes cut within the *tk* gene, *Not* I cuts within the *En-1* insert twice (see Fig 6.5). The digested samples were electrophoresed, Southern blotted and hybridised to either the *tk* oligonucleotide (oligo 383) or the *En-1* oligonucleotide (oligo 471), which hybridises to sequences 3' of the *Not* I site found in the homeo-box *En-1*. From Fig 6.6b it can be seen that with regard to the double digest each

probe hybridises to the same fragment in both plasmids, therefore in both pEN-TK1 and pEN-TK3 the 3' sequences of the *En - 1* fragment and the *tk* gene are located on the same fragment. The *En-1* insert is therefore orientated in the same direction in both pEN-TK1 and pEN-TK3 (see Fig 6.6c).

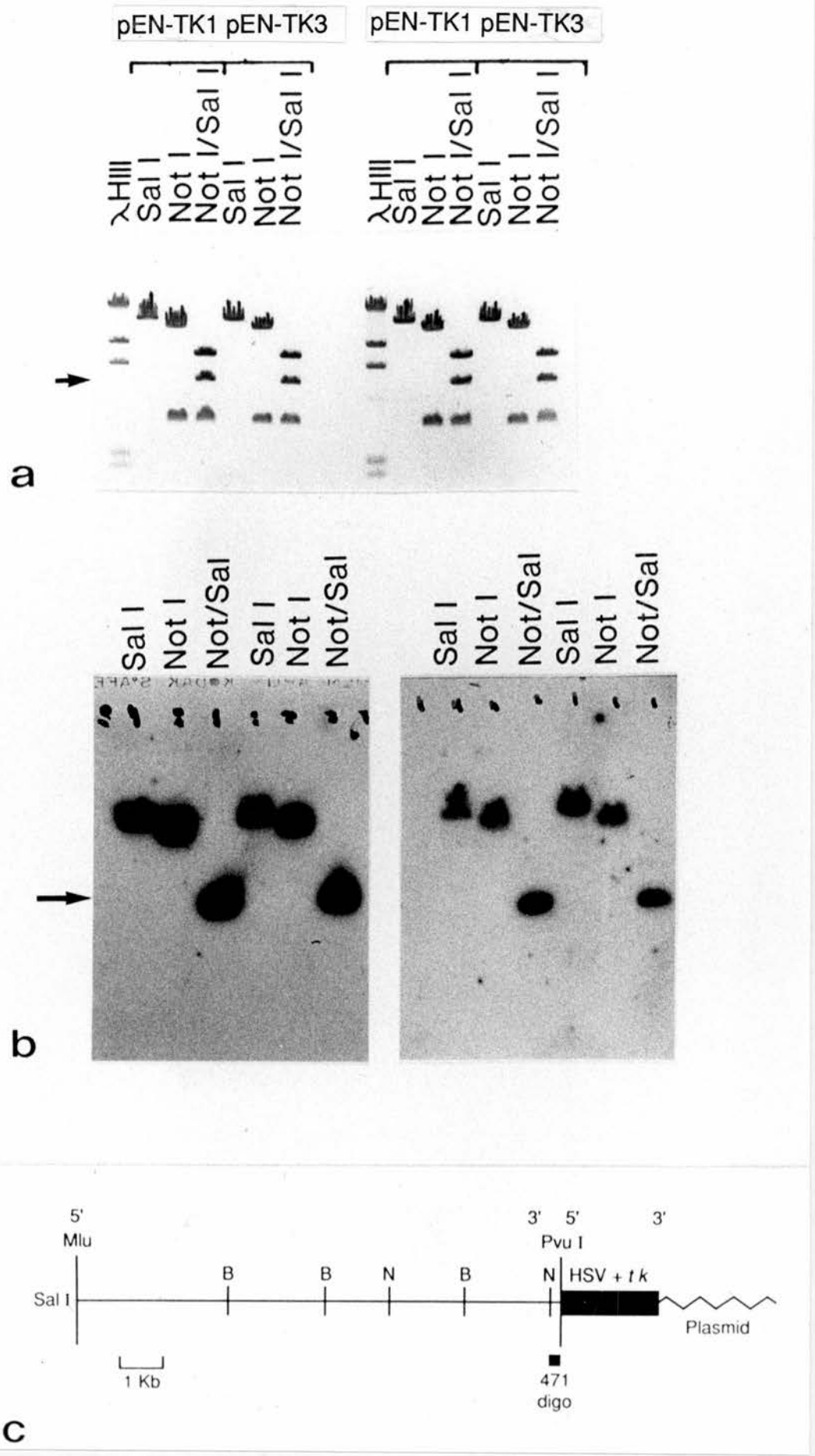
6.2.4 *Attempts to Subclone the Neo-Leu Cassette into pEN-TK1/3*

The final step in the construction of this targeting vector was the cloning of the *neo*^{r} and the yeast selectable LEU2 gene into the *Not* I sites of the cosmid fragment with the concommitant destruction of the *Not I* sites. The *neo*^{r} and the LEU2 genes were derived from pBLN3. This was obtained from Dr. Tony Brooks who had previously cloned the *Neo*^{r} and LEU2 genes into pBluescribe in a head to head orientation (see Fig 6.7). Two approaches were used in order to subclone the Neo-Leu cassette into pEN-TK1/3.

1) Blunt End Ligation

The Neo-Leu cassette can be released from pBLN3 using enzymes which resulted in blunt ends. Therefore pBLN3 was digested to completion with *Sma* I and *Nru I*. Separation of the insert from the plasmid was achieved by electrophoresis in 0.6-0.7% LMP agarose. A gel slice containing the insert was cut out and treated with Gene-Clean (see materials and methods). Insert was then cloned into pEN-TK1 which had been digested with *Not* I and then treated with T4 DNA polymerase.The ligation mix was transformed into competent JM83 and then plated out onto a 20x20cm^2 Hybond membrane. Two replica filters were made from the master plate, processed using the Grunstein & Hogness method and screened with either the *En-1* oligo 471 or with an oligo which hybridises to the *neo* ^{r} gene.

Figure 6.6 (*a*) Ethidium bromide stained gel of pEN-TK1/3 digested with *Not* I, *Sal* I or *Not* I/*Sal* I. Visible λ Hind III markers are 23kb, 9.4kb, 6.7kb, 2.3kb and 2.0kb. An arrow indicates the fragment in the *Not* I/*Sal* I double digest to which the *tk* oligo (383) and the *En-1* oligo (471) hybridise. (*b*) Southern blot analysis of the gels described in (*a*). The left hand panel was probed with the *tk* oligo (383); the right hand panel was probed with oligo 471, an *En-1* specific probe. The arrow indicates hybridising bands described above (*c*) Diagrammatic representation of pEN-TK1/3 showing the orientation of *En-1* relative to HSV-*tk*.


pEN-TK1 pEN-TK3
pEN-TK1 pEN-TK3
λHIII
Sal I
Not I
Not I/Sal I
Sal I
Not I
Not I/Sal I
a
Sal I
Not I
Not/Sal
Sal I
Not I
Not/Sal
b
5'
Mlu
Sal I
B
B
N
B
3'
5'
Pvu I
N
3'
HSV + tk
Plasmid
1 Kb
471
digo
c

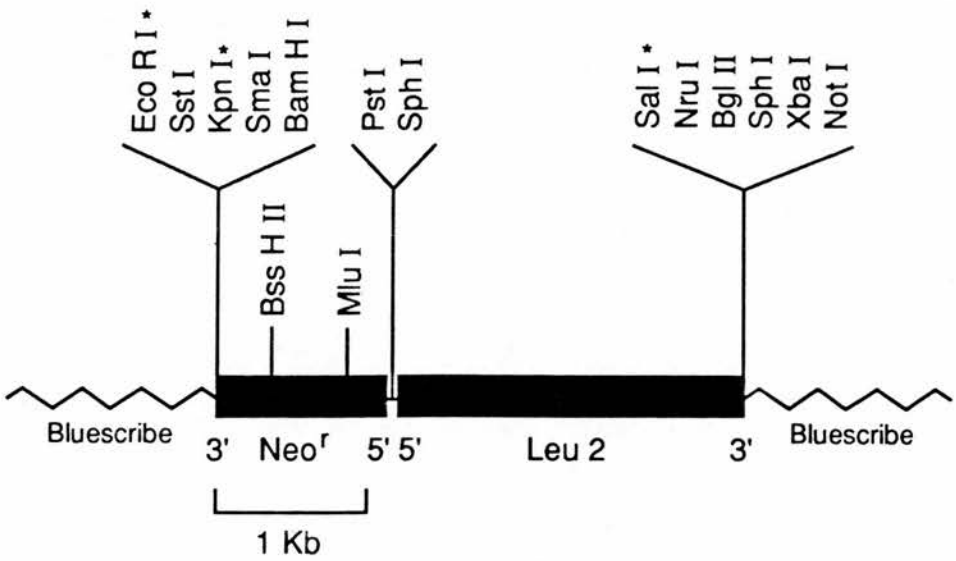


Figure 6.7 Schematic representation of the Neo-Leu cassette. The *Bss* HII and *Mlu* I sites within the *neo* gene are shown. Sites in the polylinker which are also present in the Neo-Leu cassette are denoted by an asterisk.

Several attempts at blunt end ligation failed to yield any colonies positive for both probes.

2) Cohesive End Ligation of the Neo-Leu Cassette.

As blunt end ligation failed to subclone the Neo-Leu cassette it was decided to try cohesive end ligation as this is a far more efficient reaction (Rodriguez & Tait, 1983) This was attempted by ligating synthesised linkers onto the blunt ended Neo-Leu cassette and then ligating this into the *Not* I sites of pENTK1. These linkers allow the Neo-Leu cassette to ligate into the *Not* I sites of the *En-1* fragment but do not reconstitute the *Not* I restriction sites.

2.1)Preparation of Phosphorylated Linkers

Two oligonucleotide linkers were synthesised, a 19 mer (oligo 663) and a 15 mer (oligo 662), which, when annealed together, would generate a 5' overhang (see Fig 6.8). As only phosphorylated molecules are a substrate for bacteriophage T4 DNA ligase the 662 oligo was phosphorylated before annealing to 663. A small aliquot of the annealed oligo was then electrophoresed on a 4% Nusieve gel to check that that annealing had occurred. Annealed oligo was then ethanol precipitated and resuspended in TE (10mM Tris, 1mM EDTA, pH7.4).

2.2) Preparation of the Neo-Leu Insert.

To release the Neo-Leu insert pBLN3 was digested to completion with *Sma* I and *Nru* I. The concentration of DNA was checked on a gel before ligation of the annealed linkers (see Materials & Methods). After the ligation reaction was completed

the ligase was heat inactivated (65°C, 15') and then the DNA was digested with *Xba* I and *Pvu* I. *Xba* I removes the linkers from one end of the plasmid band whereas *Pvu* I cuts the plasmid band into three bands, thus after digestion with these two enzymes only the Neo-Leu band should have the appropriate cohesive ends to ligate into the vector. This mix was then put through a 'Spun Column' in order to remove any excess linkers. The concentration of the recovered DNA was checked on an agarose gel and then aliquots were ligated to *Not* I digested pEN-TK1. The ligation mix was transformed into competent JM83 or commercially supplied DH5α cells (see Materials & Methods) and plated out onto a 20x20cm^2 Hybond membrane. Replica filters were made from the master plate, processed using the Grunstein & Hogness method and screened with either the *En-1* or the *neo*^{r} oligo. No colonies positive for both probes were found.

Linkers were also ligated onto purified Neo-Leu insert. The insert was prepared by electrophoresing digested pBLN3 on a 0.6% LMP agarose gel. The gel slice containing the Neo-Leu insert was purified using Gene-Clean.

In both instances the ligation of the linkers was checked by electrophoresis followed by Southern blotting and hybridisation to the 663 oligo.

5'G C* G G C C G C 3' *Not* I Recognition Site

3'C G C C G G* C G 5'

5'ATC AGC TCC ATA GAC 3' oligo 662

3'TAG TCG AGG TAT CTG CCG G 5' oligo 663

Fig. 6.8 Chemically synthesised oligonucleotide linkers which when annealed together as shown above generate a 5' overhang which allows ligation into a *Not* I site without reconstitution of the original *Not* I site. Altered bases of the *Not* I recognition site are indicated in bold type.

6.3 Discussion

Although the early steps in the construction of this targeting vector were relatively straight forward the final step, namely the subcloning of the Neo-Leu cassette was repeatedly unsuccessful. With regard to the attempts at blunt end ligation the problem may have been two-fold. Firstly, and the same will be true with regard to cohesive end ligation , although the overall effect of subcloning the Neo-Leu cassette will be one of replacement rather than insertion the vector has already reached a size of approximately 15 kb, which is a limiting size for plasmid vectors (Sambrook *et al,*1989). One can therefore envisage that the plasmid will tend more towards the deletion of sequences rather than their replacement. Secondly, since the efficiency of transformation of a plasmid is inversely related to the size of the plasmid (Sambrook *et al* ,1989), this together with the fact that blunt end ligation is a relatively inefficient process (Rodriguez & Tait, 1983) probably influenced the outcome of the blunt-end ligation reaction.

It was hoped that cohesive-end ligation might be a more fruitful means of subcloning the Neo-Leu cassette. However it was again disappointing that repeated attempts at ligation were unsuccessful. Although double positives were occasionally found mini prep analysis revealed that these were due to the presence of the two parental plasmids - pEN-TK1 and pBLN3. A post doctoral scientist, Dr. David Hughes, has recently begun working on this project and is attempting a slightly different strategy for subcloning the Neo-Leu cassette with the concommitant destruction of the *Not* I sites. Once this final step has been achieved gene targeting followed by YAC library construction can proceed.

As mentioned previously the vector will be introduced into RAG cells, a murine cell line which is amenable to selection with both G418 and gancyclovir (S.

Jones, pers. comm.) Southern blot analysis will determine which of the G418 and gancyclovir resistant cell lines have targeted events rather than random integrations. Those that have targeted events will then be used as a source of DNA for the construction of a YAC library. Two cloning strategies will be used to make YAC libraries: (1). Cloning from defined restriction sites and / or (2). Cloning from random fragments generated from partial digestion with a frequently cutting enzyme.

Cloning from defined restriction sites is made possible due to the long range map built up using PFGE and conventional Southern blot analysis of the cosmids. *Bss* HII releases a 800kb fragment 3' to *En - 1*. The cloning of the Neo-Leu cassette between the two *Not* I sites will still leave 5' *Bss* HII sites intact, therefore digestion of targeted DNA with *Bss* HII will only generate YAC clones 3' to *En - 1*. (n.b. Neo has internal *Bss* HII site -one can therefore only generate *Bss* HII YACs if the orientation of Neo-Leu cassette is correct, i.e., neo 5', leu 3' relative to the 5' end if the *En-1* fragment). This problem will not arise if YAC libraries are made from *Not* I digested DNA. As the genomic *Not* I sites will have been destroyed in the integrated DNA it will be possible to selectively clone an 800kb *Not* I fragment which encompasses 200kb 3' to *En - 1* and 600kb 5' to *En - 1*. In order to aid the cloning of these size fragments it should be possible to enrich for correct sized fragments on preparative PFGE (Anand *et al.*, 1989).

Alternatively a YAC library could be constructed by the standard method of partial digestion with a more frequently cutting enzyme eg *Eco* RI (Anand *et al.*, 1989; Imai & Olson, 1990) resulting in the generation of overlapping clones. In the long term it will be more advantageous to generate YAC clones using this method as it may result not only in larger clones but also in overlapping clones, which would be extremely useful in determining the localisation of the *Dh* gene. Irrespective of how the YACs around *En-1* are made they can all be selected for by transformation onto a

Leu2$^-$ background strain of yeast. This procedure should produce smaller more easily manageable libraries enriched for the area of interest.

As mentioned in Chapter 4, four animals recombinant between *Dh* and *En-1* were recovered together with a number of animals recombinant between *Emv - 17* and *Dh*. These animals have, where possible, been bred to homozygosity and the offspring will then essentially be composites of FZT and C57BL/6 DNA around the *Emv-17 - En-1* region. It is proposed that the YAC clones will be used to screen Southern blots of these recombinant animals. If any of the YACs contain a region of DNA overlapping a recombination breakpoint, one would expect to see a pattern of polymorphic restriction fragments representing a composite of both parental types of DNA. This would locate in physical terms the site of recombination. Analysis of mice recombinant between *Dh* and *En-1* would indicate the minimum distance of *Dh* relative to *En-1* whereas analysis of animals recombinant between *Dh* and *Emv-17* would determine the maximum distance within which the *Dh* gene can be located. Once this region has been localised a search for candidate genes can begin.

One method of analysis of any candidate gene is the fusion of YAC clones with embryonic stem (ES) cells (S. Jones, pers. comm.). Fusion of YACs with an intact *neo*^{r} gene could be selected for by virtue of G418 resistance (although this will not apply to any *Bss* HII generated clones see above). Pluripotent ES cells, even after they have been targeted, have the ability to populate mouse blastocysts and can contribute to the germline of the resulting chimaeras (Bradley *et al.*,1984; Thompson *et al.*,1989). Thus any YAC containing a putative *Dh* gene could be fused with ES cells and then reintroduced into the mouse blastocyst to produce a genome with three copies of the *Dh* locus. Dominant hemimelia is a semi-dominant mutation and could therefore be the result of a loss of or a gain of function. If the *Dh* phenotype is a result of loss of function then the introduction of a functional *Dh* gene into *Dh* /+ ES cells

(*Dh* /+/+ genotype) should rescue the *Dh* /+ phenotype and produce phenotypically wild type (+/+) animals. However if *Dh* is a gain of a function, the *Dh* /+/+ animals should show a *Dh* /+ phenotype. Similarily YAC clones from *Dh/Dh* and +/+ ES cells could be used to produce a +/+/*Dh* genotype.

Although the main purpose behind the design and constuction of this vector was to aid the cloning of the *Dh* gene, it can also have a secondary function.In *Drosophila* the *engrailed* gene is known to be important in segmentation, however the role of *En-1* and *En-2* in the mouse is unknown.

To elucidate the function of *En-1* in the mouse mutants must be generated. This vector can target *En-1* in ES cells and consequently disrupt the function of the *En-1* gene. As described above targeted ES cells can result in the germline transmission of the targeted gene. These mutant animals can then be bred to homozygosity allowing the mutant phenotype to be observed in both the hetero- or homozygous state. Homologous recombination into *En-2* has already been reported (Joyner *et al.*, 1989) but as yet no phenotype has been described in either the hetero- or homozygous state. That the loss of *En-2*, which is thought to establish spatial domains in the brain (Davis *et al* , 1988; Davidson *et al.*, 1988) shows no obvious phenotype is surprising, however it may be that loss of *En-2* is compensated for by the normal expression of *En-1* . Although the expression of *En-1* and *En-2* are thought to be subject to different controls their expression is not exclusive and *En-2* is in fact expressed in a smaller subset of *En-1* expressing cells (Davidson *et al.*, 1988). It would therefore be interesting to determine whether loss of *En-1* function, which has a much more diverse expression pattern, shows any phenotype. If, like *En-2* , loss of *En-1* alone shows no phenotype it would be of great interest to determine whether the combined loss of both genes would show any phenotype.

Chapter Seven

Discussion

7.1 Summary of results

In this study, all offspring were scored for *Dh* on the basis of hind limb abnormality and the presence / absence of the spleen. There was found to be complete concordance with regard to the presence of hind limb abnormality and asplenia. The limb abnormalities were, as expected, confined to the preaxial side of the limb. Although the abnormalities were extremely variable the most common abnormality was a lengthening of the hallux. The long bones of the limbs were not examined directly, but again exhibited a variable expression. The abnormalities of either the digits or the long bones were not necessarily the same on both limbs, nor was it the case that both limbs were always affected.

The number of *Dh* /+ animals recovered was not influenced by whether the gene was of maternal or paternal origin. With regard to the expressivity of the gene, although no systematic records were kept on the degree of severity, individual litters from male and female *Dh* /+ parents often included severe and mild forms of limb abnormalities. This suggests that the expressivity of *Dh* is not rigidly related to parental origin and is therefore unlikely to involve genomic imprinting.

Although *Dh* and *En-1* are closely linked on the same region of chromosome 1, they are separable by recombination (4/563), *Dh* maps 0.7cM proximal to *En-1* and 1.1cM distal to *Emv-17*. This data provides genetic evidence that *Dh* is not a mutant allele of *En-1*. The four animals recombinant between *Dh* and *En-1*, and several animals recombinant between *Dh* and *Emv-17* were, when possible, bred to homozygosity.

Short range mapping around *En-1* detected the presence of a 5' CpG island and generated two flanking probes which were used in the long range mapping.The physical localisation of *Dh* was attempted using PFGE. The PFGE map around *En-1*

encompasses 1500kb and has detected the presence of five CpG islands. There are two islands situated 3' to *En-1* and three islands situated 5' to *En-1*. Whether any of these islands are associated with the *Dh* locus depends on the ratio of genetic to physical distance in this region of the genome. As the ratio of genetic to physical distance can show a wide variation, hoping to locate the *Dh* locus simply by looking for CpG islands is futile. A means of specifically narrowing down the physical distance within which the *Dh* gene is situated was required. This was attempted by trying to localise the recombination breakpoint of each of the four mice recovered from the linkage analysis which are recombinant between *En-1* and *Dh*. Although there was some success with the technique employed for this analysis, i.e. long range mapping of partial restriction digest fragments, the recombination breakpoints were not determined.

It is vital to orientate the *En-1* gene relative to the chromosome. As *Dh* is known to map proximal to *En-1*, knowing the orientation of *En-1* would allow one to concentrate the search for the *Dh* locus in only one direction. Although the above analysis would have determined the orientation of *En-1* this was unsuccessful and therefore attempts to orientate *En-1* via the linking of the molecular probes for *En-1* and *Emv-17* were also performed. This too proved to be unsuccessful. This was probably due to the fact that, genetically, these loci are 1.8cM apart (see chapter four) which, on average, is equivalent to 3.6Mb. The probes presently available for each of these loci presumably cannot overlap this huge molecular distance. The isolation of flanking probes more distant from *En-1* will aid the orientation of *En-1*, as will localisation of the recombination breakpoints.

Several attempts, via PFGE, to detect any sizeable chromosomal rearrangements in *Dh* / *Dh* DNA proved inconclusive.

The major disadvantage of PFGE is that it does not provide any cloned DNA, and with this factor specifically in mind, the *En-1* targeting vector was designed and constructed. This vector was designed so that small easily manageable YAC libraries enriched for the *En-1* region could be generated. Resulting YAC clones can be used to (1) identify the recombination breakpoints, (2) generate flanking probes for the extension of the PFGE mapping studies and (3) locate and functionally test any putative *Dh* gene.

7.2 Future Work

The major aim of this project was to determine whether *Dh* and *En-1* were allelic and, if not, to devise an approach to localise and clone the *Dh* gene. It can be seen that this project is now at a point whereby *Dh* and *En-1* are known to be physically separate loci and the resources are available to begin localising and cloning the *Dh* gene.

The PFGE mapping is at an early stage but has already begun to identify genes which neighbour *En-1*. It is important to continue with the PFGE mapping around *En-1* and in particular to determine (1) the orientation of *En-1* and (2) the localisation of the recombination breakpoint. The importance of achieving these two goals has been discussed in chapter five.

The long term goal of this project is to clone and characterise the *Dh* gene. The *En-1* targeting vector provides the means to clone the *Dh* gene. As discussed in chapter six, numerous attempts by myself to complete the final cloning step (introduction of the Neo-Leu cassette with the concomitant destruction of the *Not* I sites) in the construction of the targeting vector were unsuccessful. The failure to complete this

final step was compounded by a time factor as this step was performed during the final weeks of my input into this project. The construct has recently been completed and targeting experiments into RAG cells are underway.

The first targeting experiments into RAG cells have been extremely encouraging (D. Hughes, K. Newton). Of approximately 300 clones which were positively selected for neomycin resistance approximately 6% (18/300) were further selected with gancyclovir. In theory therefore these 18 clones have not integrated the HSV-*tk* gene and should therefore represent targeted integrations into the *En-1* gene. It is quite possible that some of these doubly selected clones are false positives due to the complete or partial loss of the t*k* gene during integration. Based on previously published data one would expect 1-2 of these clones to contain targeted integrations (Mansour *et al.*, 1988; Johnson *et al.*, 1989). The 18 clones are in the process of being analysed via Southern blot and PCR analysis. Any clones which do contain targeted events will be used as a source of DNA for the construction of the *En-1* enriched YAC library.

YAC libraries have the capacity to clone large inserts and there have been reports of 2Mb inserts (L. Stubbs, pers. comm.). The ability to clone such large fragments greatly enhances the speed with which one can move from the *En-1* locus to *Dh*. It is possible that any resultant YAC clones will contain the *Dh* gene. If however they do not they will provide probes which can (1) extend the PFGE mapping studies and (2) be used to screen other YAC libraries. The functional analysis of YAC clones containing a putative *Dh* gene can be achieved via the fusion of the YAC clone with ES cells followed by the re-introduction of the ES cells into the mouse blastocyst. The cells derived from the ES cells in the resulting chimaeric mice will have three copies of the putative *Dh* locus. *Dh* is a semi-dominant mutation and could therefore be the result of a loss or gain of function. If *Dh* results from a loss of function then the introduction

of a functional copy of the *Dh* gene into *Dh* /+ ES cells should produce phenotypically wild type (+/+) mice with a *Dh* /+/+ phenotype. If the mutation is the result of a gain of function then the introduction of YAC clones made from *Dh*/*Dh* DNA into +/+ ES cells should produce a phenotype equivalent to *Dh* /+.

As the phenotype of *Dh* is pleiotropic it may be the result of a chromosomal deletion which, as well as deleting the *Dh* gene, is also deleting several independent genes. Alternatively the *Dh* gene may be affecting each tissue independently. The existence of *luxate* and *luxoid*, two genetically independent loci, with very similar pleiotropic phenotypes suggests that the pleiotropism seen in *Dh* animals is the result of a single gene effect. It may be that these are related genes and therefore the cloning of *Dh* may facilitate the isolation of other luxoid type genes. As discussed in the introduction to this thesis it is possible to construct a feasible chain of events to explain the pleiotropic effects of *Dh* and it was postulated that the primary effect of *Dh* occurred during somitogenesis. It is proposed that *Dh* disrupts the process of somitogenesis and this results in the loss of the normal organised structure of the splanchnic mesoderm and in vertebral malformations. The pleiotropic phenotype is a direct consequence of the combined effect of these two abnormalities.

If, as proposed, *Dh* is affecting somitogenesis it would be of interest to investigate the innervation of these animals with regard to spinal nerves. In the mouse the notocord induces the formation of the neural folds which then fuse to form the neural tube. A repetitive series of swellings or neuromeres form (Tuckett *et al.* ,1985) from which motor axons grow laterally through the anterior sclerotome towards the myotome of the adjacent somite. If extra somites are transplanted to a region on one side of the neural tube of the axolotl embryo extra spinal nerves develop (Detwiler, 1934). More recently it has been shown that a rotation of somites 180^{o} with respect to the spinal cord resulted in the motor neurones growing out from different positions so

that they still entered the original anterior half of the sclereotome (Keynes & Stern, 1984). A similar rotation of the neural tube does not affect neural segmentation. These findings suggest that the somites play a crtical role in the formation of the segmental pattern of spinal nerves.

Dh and *En-1* are now known to be physically distinct loci. It has previously been shown that *En-1* mRNA transcripts are of normal size and abundance in *Dh* /+ and *Dh/Dh* embryos (Martin *et al.*, 1990). This indicates that , certainly at the gross level , *En-1* is not responsible for the *Dh* mutation. This, however, does not rule out the possibility that *Dh* and *En-1* do have some developmental relationship. If, as proposed in the introduction *Dh* is primarily affecting the process of segmentation it is interesting to note *En-1* is expressed in the somites periodically throughout the loosely packed cells of the developing vertebral column (Davis & Joyner, 1988; Davidson *et al.*, 1988). It may be that *Dh* interferes with *En-1* expression. It would therefore be of interest to determine, via *in situ* hybridisation, the expression of *En-1* in *Dh* animals.

The *En-1* targeting vector was primarily designed and constructed with the aim of cloning the *Dh* locus. However it can also be used for determining the function of the *En-1* gene in the mouse. There are a growing number of murine genes thought to have an important developmental function (Kessel & Gruss, 1990). In order to establish what these genes are doing in the mouse mutants will have to be made. One method of doing this is to cause a loss of function of the gene, via gene targeting in ES cells (Capecchi, 1989). By knocking out one copy of the gene in ES cells followed by the reintroduction of the ES cells into a mouse blastocyst one can produce chimaeric mice. If this mutation is transmitted through the germline one can breed the resulting animals to homozygosity and therefore create homozygous null mutant mice. This technique has already been applied to a number of genes including HPRT, *int*-1, aP2

En-2 and *Hox 1.5* (Thompson *et al.*,1989; Mansour *et al.*,1988; Johnson *et al.*, 1989; Joyner *et al.*, 1991; Chisaka & Capecchi, 1991).

A mutated copy of the gene in ES cells can be produced by replacing the endogenous gene with a mutated gene via homologous recombination. The *En-1* construct contains an 11kb region of the *En-1* gene within which the neomycin resistance gene has been cloned. Heterozygous *En-1* null mutant animals can be produced by targeting the *En-1* construct directly into ES cells. Homozygous null mutant animals can be produced by breeding heterozygotes. *En-2* null mutants have already been reported (Joyner *et al.*, 1991). *En-1* and *En-2* are both expressed in the midbrain / hindbrain junction. *En-1* is also expressed in specific domains of the spinal cord, somites and limbs. In the adult, although both genes continue to show co-ordinate expression within the brain, *En-2* alone is expressed in the granule cell layers of the cerebellum (Davidson *et al.*, 1988; Davis & Joyner, 1988; Davis *et al.*, 1988). The complete absence of the *En-2* protein in the *En-2* homozgous null mutant mice has no discernible effects on viability and does not appear to have any developmental effect. The mice, however, do exhibit a subtle cerebellar dysmorphology, the cerebellum is the only tissue where *En-2* alone is expressed. As *En-1* and *En-2* are expressed in the same embryonic domains it may be that there is a functional redundancy of *En-2*. If this is the case it would be of great interest to determine the effect of the loss of *En-1* expression.

REFERENCES

Agnish, N.D. and Kochhar, D.M. (1977). The role of somites in the growth and early development of mouse limb buds. *Dev. Biol.* **56**, 174-183.

Agulnik, A.I. and Ruvinsky, A.O. (1988). Gametic imprinting: receptivity of *Fu* gene to influence of a suppressor. *Mouse News Letter.* **81**, 77-78.

Akam, M. (1987). The molecular basis of metameric pattern in the *Drosophila* embryo. *Development.* **101**, 1-22.

Alberts, B. (1983). *Molecular Biology of the Cell.* Garland Publishing, N.Y.

Anand, R. (1986). Pulsed field gel electrophoresis : a technique for fractionating large DNA molecules. *TIG.* **2**, 278-283.

Anand, R., Villasante, A. and Tyler-Smith, C. (1989). Construction of yeast artificial chromosome libraries with large inserts using fractionation by pulsed field gel electrophoresis. *Nuc. Acid Res.* **17**, 3425-3433.

Antequera, F., Boyes, J. and Bird, A. (1990). High levels of *de novo* methylation and altered chromatin structure at CpG islands in cell lines. *Cell.* **62**, 503-514.

Asashima, M., Nakano, H., Shimada, K., Kinoshita, K., Ishii, K., Shibai, H. and Ueno, H. (1990). Mesodermal induction in early amphibian embryos by Activin A (erythroid differentiation factor). *Roux's Arch. Dev. Biol.* **198,** 330-335.

Bailey, N.T.J. (1959). *Statistical Methods in Biology.* English Universities Press.

Baker, N.E. (1987). Molecular cloning of sequences from *wingless*, a segment polarity gene in *Drosophila* : The spatial disribution of a transcript in embryos. *EMBO J.* **6**, 1765-1773.

Balinsky (1975). *An introduction to embryology* (4th edition). W.B. Saunders Co.

Barlow, D.P. and Lehrach, H. (1987). Genetics by gel electrophoresis: the impact of pulsed field gel electrophoresis on mammalian genetics. *TIG.* **3**, 167-171.

Barsh, G.S. & Epstein, C.J. (1989). The long range restriction map surrounding the mouse *agouti* locus reveals a disparity between physical and genetic distances. *Genomics.* **5**, 9-18.

Battisto, J.R., Borek, F. & Busci, R.A. (1971). Splenic determination of immunocompetence: Influence on other lymphoid organs. *Cell. Immunol.* **2**, 627-633.

Battisto, J.R., Cantor, L.C., Borek, F., Goldstein, A.L. and Cabrerra, E. (1969). Immunoglobulin synthesis in hereditarily spleenless mice. *Nature.* **222,** 1196-1198.

Beachy, P.A., Helfand, S.L. and Hogness, D.S. (1985). Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature.* **313**, 545-551.

Beachy, P.A., Krasnow, M.A., Gavis, E.R. and Hogness, D.S. (1988). An *Ultrabithorax* protein binds sequences near it's own and the *Antennapedia* P1 promoters. *Cell.* **55,** 1069-1081.

Bender, W., Akam, M., Karch, F., Beachy, P.A., Sperier, P., Lewis, E.B. & Hogness, D.S. (1983). Molecular genetics of the *bithorax* complex of *Drosophila melanogaster. Science.* **221**, 23-29.

Bernards, A., Kooter, J.M., Michels, P.A.M., Moberts, R.M.P. and Borst, P. (1986). Pulsed field gradient electrophoresis of DNA digested in agorase allows the sizing of the large duplication of a surface antigen in *trypanosomes. Gene.* **42**, 313-322.

Bickmore, W.A., Porteous, D.J., Christie, S., Seawright, A., Fletcher, J.M., Maule, J.C., Couillin, P., Junien, C., Hastie, N.D. and van Heyningen, V. (1989). CpG islands surround a DNA segment located between translocation breakpoints associated with genitourinary dysplasia and aniridia. *Genomics.* **5**, 685-693.

Bienz, M. and Tremml, G. (1988). Domain of *Ultrabithorax* expression in *Drosophila* visceral mesoderm from autoregulation and exclusion. *Nature.* **333**, 576-578.

Biggan, M. and Tijian, R. (1989). A purified *Drosophila* homeodomain protein represses transcription *in vitro. Cell.* **58**, 433-440.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. *Nature.* **321**, 209-213.

Bird, A.P. (1987). Two classes of observed frequency for rare-cutter sites in CpG islands. *Nuc. Acid Res.* **17**, 9485.

Bird, A.P., Taggart, M., Frommer, M., Miller, O.J. and McLeod, D. (1985). A fraction of the mouse genome that is derived from islands of non-methylated CpG rich DNA. *Cell.* **40**, 91-99.

Bird, A.P., Taggart, M.H., Nicholls, R.D. and Higgs, D.R. (1987). Non-methylated CpG rich islands at the human α-globin locus: implications for evolution of the α-globin pseudogene. *EMBO J.* **6**, 999-1004.

Bodner, M., Castrillo, J-L., Theill, L.E., Deerinck, T., Ellisman, M. and Karin, M. (1988). The pituatury specific transcription factor GHF-1 is a homeobox-containing protein. *Cell.* **55**, 505-518.

Bolivar, F., Rodrigues, R.L., Greene, R.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977). Construction and characterisation of new cloning vehicles. II. A multipurpose cloning system. *Gene.* **2**, 95-113.

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. and Noll, M. (1986). Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes of *Drosophila* . *Cell.* **47**, 1033-1040.

Boterenbrood, E.C. and Nieuwkoop, P.D. (1973). The formation of the mesoderm in urodelean amphibians. V. Its regional induction by the endoderm. *Wilhelm Roux' Archiv.* **173**, 319-332.

Botstein, D., Falco, S.C., Stewart, S.E., Brennan, M., Scherer, S., Stinchcomb, D.T., Struhl, K. and Davis, R.W. (1979). Sterile Host Yeasts (SHY) : A eukaryotic system of biological containment for recombinant DNA experiments. *Gene.* **8**, 17-24.

Boulet, A.M. and Scott, M.P. (1988). Control elements of the P2 promoter of the *antennapedia* gene. *Genes Dev.* **2**, 1600-1614.

Bradley, A., Evans, M., Kaufman, M.H. and Robertson, E. (1984). Formation of germ-line chimaeras from embryo derived teratocarcinoma cell lines. *Nature.* **309**, 255-256.

Brennan, R.G. and Mattews, B.W. (1989). Structural basis of DNA-protein recognition. *TIBS.* **14**, 286-290.

Brown, W.R.A. and Bird, A.P. (1986). Long-range restriction site mapping of mammalian genomic DNA. *Nature.* **322**, 477-481.

Buchberg, A.M., Taylor, B.A., Jenkins, N.A. and Copeland, N.G. (1986). Chromosomal localisation of *EMV-16* and *EMV-17*, two closely linked ecotropic proviruses of RF/J mice. *J. Virol.* **60**, 1175-1178.

Burke, D.T., Carle, G.F. and Olsen, M.V. (1987). Cloning large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science.* **236**, 806-812.

Capecchi, M.R. (1989). Altering the genome by homologous recombination. *Science.* **244**, 1288-1292.

Carroll, S.B. and Scott, M.P. (1985). Localisation of *fushi-tarazu* protein during *Drosophila* development. *Cell.* **43**, 47-57.

Carter, T.C. (1951). The genetics of *luxate* mice. I. Morphological abnormalities of heterozygotes and homozygotes. *J. Genet.* **50**, 277-299.

Carter, T.C. (1953). The genetics of *luxate* mice. III. Horse-shoe kidney, hydronephrosis and lumbar reduction. *J. Genet.* **51**, 441-457.

Carter, T.C. (1954). The genetics of *luxate* mice. IV. Embryology. *J. Genet.* **52**, 1-35.

Chisaka, O. & Capecchi, M.R. (1991). Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *Hox 1.5*. *Nature.* **350**, 473-479.

Cho, K.W.Y. and DeRobertis, E.M. (1990). Differential activation of *Xenopus* homeo-box genes by mesoderm inducing growth factors and retinoic acid. *Genes Dev.* **4**, 1910-1916.

Coleman, K.G., Poole, S.J., Weir, M.P., Soeller, W.C. and Kornberg, T. (1987). The *invected* gene of *Drosophila* : sequence ananlysis and expression studies reveal a close kinship to the *engrailed* gene. *Genes Dev.* **1**, 19-28.

Cote, S., Preiss, A., Haller, J., Schuh, R., Kiencin, A., Seifer, E. and Jackle, H. (1987). The *gooseberry* zipper region of *Drosophila*: five genes encode different spatially restricted transcripts in the embryo. *EMBO J.* **6**, 2793-2801.

Coulson, A., Waterston, R., Kiff, J., Sulston, J. and Kohara, Y. (1988). Genome linking with yeast artificial chromosomes. *Nature.* **335**, 184-186.

Dale, L. and Slack, J.M.W. (1987a). Fate map for the 32-cell stage of *Xenopus laevis*. *Development.* **99**, 527-551.

Dale, L. and Slack, J.M.W. (1987b). Regional specification within the mesoderm of early embryos of *Xenopus laevis*. *Development.* **100**, 279-295.

Dale, L., Smith, J.C. and Slack, J.M.W. (1985). Mesoderm induction in *Xenopus laevis:* A quantitative study using a cell lineage label and tissue specific antibodies. *J. Embryol. exp. Morphol.* **89**, 289-312.

Davidson, D., Graham, E., Sime, C. and Hill, R. (1988). A gene with sequence similarity to *Drosophila engrailed* is expressed during the development of the neural tube and vertebrae in the mouse. *Development.* **104**, 305-316.

Davis, C.A., Holmyard, D.P., Millen, K.J. and Joyner, A.L. (1991). Examining pattern formation in mouse, chicken and frog embryos with an *En* specific antisera. *Development.* **111**, 287-298.

Davis, C.A. and Joyner, A.L. (1988). Expression patterns of the homeobox containing genes *En-1* and *En-2* and the proto-oncogene *int-1* diverge during development. *Genes Dev.* **2**, 1736-1744.

Davis, C.A., Noble-Topham, S.E., Rossant, J. and Joyner, A.L. (1988). Expression of the homeobox containing gene *En-2* delineates a specific region of the developing mouse brain. *Genes Dev.* **2**, 361-371.

Desplan, C., Theis, J. and O'Farrell, P.H. (1985). The *Drosophila* developmental gene, *engrailed* , encodes a sequence specific DNA binding activity. *Nature.* **318**, 630-635.

Desplan, C., Theis, J. and O'Farrell, P.H. (1988). The sequence specificty of homeodomain DNA interaction. *Cell.* **54**, 1081-1090.

Detwiler, S.R. (1934). An experimental study of spinal nerve segmentation in *amblystoma* with reference to the plurisegmental contribution to the brachial plexus. *J. Exp. Zool.* **67**, 395-441.

Deutsch, U., Dressler, G.R. and Gruss, P. (1988). *Pax-1*, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. *Cell.* **53**, 617-625.

Dickie, M.M. (1968). *Mouse News Letter.* **38**, 24.

DiNardo, S., Kuner, J.M., Theis, J. and O'Farrell, P.H. (1985). Development of embryonic pattern in *Drosophila melanogaster* as revealed by accumulation of the nuclear *engrailed* protein. *Cell.* **43,** 59-69.

Drees, B., Ali, Z., Soeller, W., Coleman, K., Poole, S. and Kornberg, T. (1987). The transcription unit of the D*rosophila engrailed* locus : an unusually small portion of a 70,000 bp gene. *EMBO J.* **6**, 2803-2809.

Dressler, D.G. and Gruss, P. (1988). Do multigene families regulate vertebrate development ? *TIG.* **4**, 214-219.

Driever, W. and Nusslein-Volhard, C. (1988a). A gradient of *bicoid* protein in *Drosophila* embryos. *Cell.* **54,** 83-93.

Driever, W. and Nusslein-Volhard, C. (1988b). The *bicoid* protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell.* **54,** 95-104.

Driever, W. and Nusslein-Volhard, C. (1989). The *bicoid* protein is a positive regulator of *hunchback* transcription in the early *Drosophila* embryo. *Nature.* **337,** 138-143.

Duboule, D., Boncinelli, E., DeRobertis, E., Featherstone, M., Lonai, P., Oliver, G. and Ruddle, F. (1990). An update of mouse and human HOX gene nomenclature. *Genomics.* **7**, 458-459.

Duboule, D. and Dolle, P. (1989). The structural and functional organisation of the murine HOX gene family resembles that of *Drosophila* homeotic genes. *EMBO J.* **8**, 1497-1505.

Dulcibella, T. (1977). Surface changes of the developing trophoblast. In *Development in Mammals* vol. 1 (ed. M.H. Johnson). Elsevier, Amsterdam, North Holland.

Dunn, L.C. and Gluecksohn-Schoenheimer, S. (1947). A new complex of hereditary abnormalities in the house mouse. *J. Exp. Zool.* **104**, 25-51.

Dunn, L.C., Gluecksohn-Schoenheimer, S. and Bryson, V. (1940). A new mutation in the mouse affecting spinal column and urogenital system. *J. Exp. Zool.* **31**, 343-348.

Eaton, G.J. and Green, M.M. (1963). Giant cell differentiation and lethality of homozygous *yellow* mouse embryos. *Genetica.* **34**, 155-161.

Eberlein, S. and Russell, M. (1983). Effects of deficiences in the *engrailed* region of *Drosophila melanogaster*. *Dev. Biol.* **100**, 227-237.

Ehrich, E., Craig, A., Poustka, A., Frischauf, A-M. and Lehrach, H. (1987). A family of cosmid vectors with the multicopy R6K replication origin. *Gene.* **57**, 229-237.

Eker, R. (1929). The recessive mutant *engrailed* in *Drosophila melanogaster*. *Hereditas.* **12**, 217-222.

Fjose, A., McGinnis, W.J. and Gehring, W.J. (1985). Isolation of a homeo-box containing gene from the *engrailed* region of *Drosophila* and the spatial distribution of its transcript. *Nature.* **313**, 284-289.

Fletcher, M.P., Ikeda, R.M. and Gershwin, M.E. (1977). Splenic influence of T cell function : the immunobiology of the inbred hereditarily asplenic mouse. *J. Immunol.* **119**, 110-117.

Flint, O.P. (1977). Cell interactions in the developing axial skeleton in normal and mutant mouse embryos. In *Vertebrate Limb and Somite Morphogenesis* (ed. D.A. Ede, J.R. Hinchcliffe, M. Balls). Cambridge University Press, Cambridge.

Flint,O.P., Ede, D.A., Wilby, O.K., Proctor, J. (1978). Control of somite number in normal and amputated mouse embryos. *J. Embryol. exp. Morphol.* **45**, 189-202.

Folger, K.R., Wong, E.A., Wahl, G. and Capecchi, M.R. (1982). Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for

homologous recombination between injected plasmid DNA molecules. *Molec. cell. Biol.* **5**, 1372-1387.

Forsthoefel, P.F. (1958). The skeletal effects of the *luxoid* gene in the mouse, including its interaction with the *luxate* gene. *J. Morph.* **102**, 247-288.

Forsthoefel, P.F. (1959). The embryological development of the skeletal effects of the *luxoid* gene in the mouse, including its interaction with the *luxate* gene. *J. Morph.* **104**, 89-142.

Freye, H. (1954). Anatomische und entwicklungsgeshichtliche Untersuchungen am Skelett normale und oligodactyler Mause. *Wiss. Z. Univ. Hall, Math-Nat.* **3**, 801-824.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. and Noll, M. (1986). Structure of the segmentation gene *paired* and the *Drosophila* PRD gene set as part of a gene network. *Cell.* **47**,735-746.

Frohnhofer, H.G. and Nusslein-Volhard, C. (1986). Organisation of anterior pattern in the *Drosophila* embryo by the maternal gene *bicoid. Nature.* **324**, 120-125.

Frohnhofer, H.G. and Nusslein-Volhard, C. (1987). Maternal genes required for the anterior localisation of *bicoid* activity in the embryo of the *Drosophila. Genes Dev.* **1**, 880-890.

Garcia-Bellido, A., Ripoll, P. and Morata, G. (1973). Developmental compartmentalisation of the wing disc of *Drosophila. Nature New Biol.* **245**, 251-253.

Gardiner-Garden, M. and Frommer, M. (1987). CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261-282.

Gaunt, S.J., Sharpe, P.T. and Duboule, D. (1988). Spatially restricted domains of homeogene transcripts in mouse embryos : relation to a segmented body plan. *Development* (Suppl.). **104**, 169-179.

Gehring, W.J. (1985). Homeotic genes, the homeobox and genetic control of development. *CSH Symp. Quant. Biol.* **50**, 243-251.

Gehring, W.J. (1987). Homeoboxes in the study of development. *Science*. **236**, 1245-1251.

Gerhart, J.C., Black, S., Scharf, S., Gimlich, R., Vincent, J-P., Danilchik, M., Rowning, B. and Roberts, J. (1986). Amphibian early development. Bioscience. **36**, 541-549.

Gerhart, J.C., Vincent, J-P., Scharf, S.R., Black, S.D., Gimlich, R.L. and Danilchik, M. (1984). Localisation and induction in early development of *Xenopus*. *Phil. Trans. R. Soc. Lond.* **307**, 319-330.

Gershwin, M.E., Ahmed, R.M., Ikeda, R.M., Shifrine, M. and Wilson, F. (1978). Immunobiology of congenitally athymic-asplenic mice. *Immunology*. **34**, 631-642.

Gershwin, M.E., Castles, J.J., Ikeda, R.M., Erikson, K. and Montero, J. (1979). Studies on congenitally immunologic mutant New Zealand mice. I. Autoimmune features of hereditarily asplenic (*Dh* /+) NZB mice, reduction of naturally occuring thymocytotoxic antibody and normal suppressor function. *J. Immunol.* **122**, 710-717.

Gimlich, R.L. and Cooke, J. (1983). Cell lineage and the induction of second nervous systems in amphibia development. *Nature*. **306**, 471-473.

Godsave, S.F., Isaacs, H.V. and Slack, J.M.W. (1988). Mesoderm inducing factors: a small class of molecules. *Development*. **102**, 555-556.

Graham, A., Papalopulu, N. and Krumlauf, R. (1989). The murine and *Drosophila* homeobox gene complexes have common features of organisation and expression. *Cell*. **57**, 367-378.

Green, M.C. (1955). Luxoid-A new hereditary leg and foot abnormality. *J. Hered.* **46**, 91-99.

Green, M.C. (1957). Modifiers of the pleiotropic effects of the short ear gene in the mouse. *J. Hered.* **48**, 205-209.

Green, M.C. (1967). A defect of the splanchnic mesoderm caused by the mutant gene Dominant hemimelia in the mouse. *Dev Biol.* **15,** 62-89.

Green, M.C. (1989). In *Genetic variants and strains of the laboratory mouse* (eds. M.F. Lyon & A.G. Searle). Oxford University Press, Oxford.

Gruneberg, H. (1963). *The Pathology of Development*. Blackwell Press, Oxford.

Grunz, H. and Tacke, L. (1986). The inducing capacity of the presumptive endoderm of *Xenopus laevis* studied by transfilter experiments. *Wilhelm Roux's Arch. devl. Biol.* **195,** 467-473.

Gubb, D. (1985). Further studies on *engrailed* mutants in *Drosophila melanogaster*. *Wilhelm Roux's Arch. devl. Biol.* **194,** 236-246.

Gurdon, J.B. (1987). Embryonic induction: molecular prospects. *Development*. **99,** 285-306.

Gurdon, J.B. (1988). A community effect in animal development. *Nature*. **336**, 772-774.

Gurdon, J.B. (1989). The localisation of an inductive response. *Development*. **105**, 27-33.

Hafen, E., Levine, M. and Gehring, W.J. (1984a). Regulation of *Antennapedia* transcript distribution by the bithorax complex in *Drosophila*. *Nature*. **307**, 287-289.

Hafen, E., Kuriowa, A. and Gehring, W.J. (1984b). Spatial distribution of transcripts from the segmentation gene *fushi-tarazu* during *Drosophila* embryonic development. *Cell*. **37**, 833-841.

Hamburger, V. (1988). *The heritage of experimental biology. Hans Spemann and the organiser*. Oxford University Press.

Hanahan, D. (1983). Studies on transformation of *E. coli* with plasmids. *J. Mol. Biol*. **166**, 557-580.

Hanahan, D. (1985). Techniques for the transformation of E. coli. In *DNA Cloning: a practical volume* Vol. 1. (ed. D.M. Glover). IRL Press, Oxford.

Hanes, S.D. and Brent, R. (1989). DNA specificty of the *bicoid* activator protein is determined by homeodomain recognition helix residue 9. *Cell*. **57**, 1275-1283.

Harding, K., Wedeen, C., McGinnis, W. and Levine, M. (1985). Spatially regulated expression of homeotic genes in *Drosophila*. *Science*. **229**, 1236-1242.

Hayashi, S. and Scott, P. (1990). What determines the specificity of action of *Drosophila* homeodomain proteins? *Cell*. **63,** 883-894.

Hill, R.E., Hall, A.E., Sime, C.M. and Hastie, N.D. (1987). A mouse homeo box-containing gene maps near a developmental mutation. *Cytogenet. Cell Genet.* **44**, 171-174.

Hiromi, Y. and Gehring, W.J. (1987). Regulation and function of the *Drosophila* segmentation gene *fushi-tarazu*. *Cell*. **50,** 963-974.

Hoey, T. and Levine, M. (1988). Divergent homeo-box proteins recognise similar DNA sequences in *Drosophila*. *Nature*. **332**, 858-861.

Hogan, B.H., Costantini, F. and Lacy, E. (1986). *Manipulating the Mouse Embryo, A Laboratory Manual*. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.

Holland, P.W.H. & Hogan, B.L.M. (1988). Expression of homeobox genes during mouse development: a review. *Genes & Development*. **2**, 773-782.

Howard, K. and Ingham, P. (1986). Regulatory interactions between the segmentation genes *fushi-tarazu, hairy* and *engrailed* in the *Drosophila* blastoderm. *Cell*. **44,** 949-957.

Imai, T. and Olson, M.V. (1990). Second-generation approaches to the construction of Yeast Artificial Chromosome Libraries. *Genomics*. **8**, 297-303.

Ingham, P., Martinez-Arias, A., Lawrence, P. and Howard, K. (1985). Expression of *engrailed* in the parasegment of *Drosophila*. *Nature*. **317**, 634-636.

Ingraham, H.A., Chen, R., Mangalam, H.J., Elshotz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. and Rosenfeld, M.G. (1988). A tissue specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell*. **55**, 519-529.

Jacobsen, A.G. (1966). Inductive process in embryonic development. *Science*. **152**, 25-34.

Jacobsen, A.G. and Slater, A.K. (1988). Features of embryonic induction. *Development.* **104**, 341-359.

Jenkins, N.A., Copeland, N.G., Taylor, B.A., Bedigan, H.G. and Lee, B.K. (1982). Ecotropic murine leukemia virus DNA content of normal and lymphomatous tissue of BxH2 recombinant inbred mice. *J. Virol.* **42**, 379-388.

Johnson, R.S., Sheng, M., Greenberg, M.E., Kolodner, R.D., Papaioannou, V.E. & Spiegelman, B.M. (1989). Targeting of nonexpressed genes in embryonic stem cells via homologous recombination. *Science.* **245**, 1234-1236.

Johnson, M.H. and Ziomek, C.A. (1981). The foundation of two distinct cell lineages within the mouse morula. *Cell.* **24**, 71-80.

Joyner, A.L., Herrup, K., Auerbach, B.A., Davis, C.A. and Rossant, J. (1991). Subtle cerebellar phenotype in mice homozygous for a targeted deletion in the *En-2* homeobox. *Science.* **251**, 1239-1242.

Joyner, A.L., Kornberg, T., Coleman, K.G., Cox, D.R. and Martin, G.R. (1985). Expression during embryogenesis of a mouse gene with sequence homology to the *Drosophila engrailed* gene. *Cell.* **43**, 29-37.

Joyner, A.L. and Martin, G.R. (1987). *En-1* and *En-2*, two mouse genes with sequence homolgy to the *Drosophila engrailed* gene: expression during embryogenesis. *Genes Dev.* **1**, 29-38.

Joyner, A.L., Skarnes, W.C. and Rossant, J. (1989). Production of a mutation in the mouse *En-2* gene by homologous recombination in embryonic stem cells. *Nature.* **338**, 153-156.

Kao, K.R., Masui, Y. & Elinson, R.P. (1986). Lithium induced respecification of pattern in *Xenopus laevis* embryos. *Nature.* **322**,371-373.

Karr, T.L., Ali, Z., Drees, B. and Kornberg, T. (1985). The *engrailed* locus of *Drosophila melanogaster* provides an essential zygotic function in precellular embryos. *Cell.* **43**, 591-601.

Kaufman, T.C., Lewis, R. and Wakimoto, B. (1980). Cytogenetic analysis of chromosome three in *Drosophila melanogaster*: the homeotic gene complex in polytene chromosome interval 84A-B. *Genetics*. **94**, 115-133.

Kelly, S.J. (1977). Studies on the development potential of 4- and 8-cell stage mouse blastomeres. *J. Exp. Zool.* **200**, 365-376.

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

Kessel, M., Schulze, F., Fibi, P. and Gruss, P. (1987). Primary structure and nuclear localisation of a murine homeodomain protein. *Proc. Natl. Acad. Sci.* (USA). **84**, 5306-5310.

Keynes, R.J. and Stern, C.D. (1984). Segmentation in the vertebrate nervous system. *Nature*. **310**, 786-789.

Kieny, M. (1971). Les phases d'activite morphogenes du mesoderme somatopleural pendant le development precoce du membre chez l'embryon de poulet. *Annales d'Embryologie et du Morphogenese*. **4**, 281-288.

Kieny, M., Manger, A. and Senger, P. (1972). Early regionalisation of the somitic mesoderm as studied by the development of the axial skeleton of the chick embryo. *Dev. Biol.* **28**, 142-161.

Kimmelman, D., Abraham, J.A., Haaparanta, T., Palisi, T.M. and Kirschner, M.W. (1988). The presence of fibroblast growth factor in the frog egg: It's role as a natural mesoderm inducer. *Science*. **242**, 1053-1056.

Kimmelman, D. and Kirschner, M.W. (1987). Synergistic induction of mesoderm by FGF and TGF-β and the identification of a mRNA coding for FGF in the early *Xenopus* embryo. *Cell*. **51**, 869-877.

Kintner, C.R. and Melton, D.A. (1987). Expression of *Xenopus* N-CAM RNA in ectoderm as an early response to neural induction. *Development*. **99**, 311-320.

Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990). Crystal structure of an *engrailed* homeodomain-DNA complex at 2.8 Å

resolution: A framework for understanding homeodomain-DNA interactions. *Cell.* **63**, 579-590.

Knudsen, T.B. and Kochhar, D.M. (1981). The role of morphogenetic cell death during abnormal limb-bud outgrowth in mice heterozygous for the dominant mutation *Hemimelic extra toe* (*Hm*^{x}). *J. Embryol. exp. Morphol.* (Suppl.). **65**, 289-307.

Ko, H.S., Fast, P., McBride, W. and Staudt, L.M. (1988). A human protein specific for the immnuoglobulin octamer DNA motif contains a functional homeobox domain. *Cell.* **55**, 135-144.

Kornberg, T. (1981a). *engrailed*: A gene controlloing compartment and segment formation in *Drosophila. Proc. Natl. Acad. Sci.* (USA). **78**, 1095-1099.

Kornberg, T. (1981b). Compartments in the abdomen of *Drosophila* and the role of the *engrailed* locus. *Dev. Biol.* **86**, 363-372.

Kornberg, T., Siden, I., O'Farrell, P.H. and Simon, M. (1985). The *engrailed* locus in *Drosophila*: in situ localisation reveals compartment specific expression. *Cell.* **40**, 45-53.

Kornfeld, K., Saint, R.B., Beachy, P.A., Harte, P.J., Peattie, D.A. and Hogness, D.A. (1989). Structure and expression of a family of *Ultrabithorax* mRNAs generated by alternative splicing and polyadenylation in *Drosophila. Genes Dev.* **3**, 243-258.

Krasnow, M.A., Saffman, E.E., Kornfeld, K. and Hogness, D.S. (1989). Transcriptional activation and repression by *Ultrabithorax* proteins in cultured *Drosophila* cells. *Cell.* **57**, 1031-1043.

Kuner, J., Nakanishi, M., Ali, Z., Drees, B., Gustavson, E., Theis, J., Kauvar, L.M., O'Farrell, P. and Kornberg, T. (1985). Molecular cloning of *engrailed*: A gene involved in the development of pattern in *Drosophila melanogaster. Cell.* **42**, 309-316.

Kuriowa, A., Hafen, E. and Gehring, W.G. (1984). Cloning and transcriptional analysis of the segmentation gene *fushi-tarazu* of *Drosophila melanogaster. Cell.* **37**,825-831.

Kuziora, M.A. & McGinnis, W. (1989). A homeodomain substitution changes the regulatory specificity of the *Deformed* protein in *Drosophila* Embryos. *Cell.* **59**, 563-571.

Laughon, A., Howell, W. and Scott, M.P. (1988). The interaction of proteins encoded by *Drosophila* homeotic and segmentation genes with specific DNA sequences. *Development* (Suppl). **104**, 75-83.

Laughon, A. and Scott, M.P. (1984). Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. *Nature.* **310**, 25-31.

Lawrence, P. and Morata, G. (1976). Compartments in the wing disc of *Drosophila*: a study of the *engrailed* gene. *Dev. Biol.* **50**, 321-337.

Lawrence, S.K., Smith, C.L., Srivastava, R., Cantor, C.R. and Weissman, S.M. (1987). Megabase-scale mapping of the HLA gene complex by pulsed field electrophoresis. *Science.* **235**, 1387-1390.

Lawrence, P. and Struhl, G. (1982). Further studies of the *engrailed* phenotype in *Drosophila. EMBO. J.* **1**, 827-833.

Lawrence, P., Struhl, G. and Morata, G. (1979). Bristle patterns and compartment boundaries in the talsi of *Drosophila. J. Embryol. exp. Morphol.* **51**, 195-208.

Lehmann, R. and Nusslein-Volhard, C. (1987). *hunchback*, a gene required for segmentation in the anterior and posterior regions of the *Drosophila* embryo. *Dev. Biol.* **119**, 402-417.

Levine, M., Rubin, G.M. & Tijian, R. (1984). Human DNA sequences homologous to a protein coding region conserved between homeotic genes of *Drosophila. Cell.* **38**,667-673.

Lewis, R.H. (1904). Experimental studies on the eye in Amphibia. I. On the origin of the lens in *Rana palustris. Amer. J. Anat.* **3**, 505-536.

Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila. Nature.* **276**, 565-576.

Lewis, W.H. and Wright, E.S. (1935). On the early development of the mouse egg. *Carnegie Inst. Wash. Pub.* **459**, 133-144.

Lindsay, S. and Bird, A.P. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature*. **327**, 336-338.

Lozzio, B.B. (1972). Hematopoiesis in congenitally asplenic mice. *Am. J. Physiol.* **222**, 290

Lozzio, B.B. & Machado, E.A. (1975). Influence of a neonatal spleen transplant on hematopoiesis of hereditarily asplenic mice. *Exp. Hematol.* **3**, 156-168.

Lozzio, B.B. and Wargon, L.B. (1974). Immune competence of hereditarily spleenless mice. *Immunology*. **27**, 167-178.

Lyon, M.F. & Searle, A.G. (1989). *Genetic variants and strains of the Laboratory mouse*. Oxford University Press, Oxford.

Macdonald, P.M., Ingham, P. and Struhl, G. (1986). Isolation, structure and expression of *even-skipped*: a second pair rule gene of *Drosophila* containing a homeo-box. *Cell*. **47**, 721-734.

Macdonald, P.M. and Struhl, G. (1986). A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. *Nature*. **324**, 537-545.

Machado, E.A. and Lozzio, B.B. (1976). Animal model: hereditary asplenia in mice. *Am. J. Pathol.* **85**, 515-518.

Malicki, J., Schugart, K. and McGinnis, W. (1990). Mouse *Hox 2.2* specificies thoraic segmental identity in *Drosophila* embryos and larvae. *Cell*. **63**, 961-967.

Mangold, O. (1933). Uber die induktions - Fahigkeit der verschiedened Bezirke der Neurula von Urodelen. *Naturwissenschaften*. **21**, 761-766.

Mangold, O. and Spemann, H. (1927). Uber induktion von Medullarplatte durch Medullarplatte im jungeren Keim, ein Beispiel homoogenetischer oder assimilaturischer induktion. *Wilhelm Roux' Arch. EntwMech. Org*. **111**, 341-422.

Mansour, S.L., Thomas, K.R. and Capecchi, M.R. (1988). Distruption of the proto-oncogene *int-2* in mouse embryo derived stem cells: a general stategy for targeting mutations to non-selectable genes. *Nature*. **336**, 348-352.

Martin, G.R. (1980). Teratocarcinomas and mammalian embryogenesis. *Science.* **209**, 768-776.

Martin, G.R., Richman, M., Reinsch, S., Nadeau, J.H. and Joyner, A.L. (1990). Mapping of two mouse engrailed-like genes: Close linkage of *En-1* to *dominant hemimelia* (*Dh*) on chromosome 1 and of *En-2* to *hemimelic extra-toes* (*Hx*) on chromosome 5. *Genomics.* **6**, 302-308.

Mason, A.J., Berkenmeier, L., Schmelzer, C. and Schwall, R. (1989). Activin B: precursor sequences, genomic structure and *in vitro* activities. *Mol. Endocrinol.* **3**, 1352-1358.

McGinnis, N., Kuziora, M.A. and McGinnis, W. (1990). Human *Hox4.2* and *Drosophila Deformed* encode similar regulatory specificities in *Drosophila* embryos and larvae. *Cell.* **43**, 969-976.

McGinnis, W., Levine, M.S., Hafen, E., Kuriowa, A. and Gehring, W.J. (1984). A conserved DNA sequence in homeotic genes of the *Drosophila Antennapedia* and *bithorax* complexes. *Nature.* **308**,428-433.

McLaren, A. (1976). Genetics of the early mouse embryo. *Ann. Rev. Genet.* **10**, 361-368.

Mead, D.A., Szczesna-Skorupa, E. and Kemper, B. (1986). Single stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. *Protein Engineering.* **1**, 67-74.

Meier, S. (1979). Development of the chick mesoblast: formation of the embryonic axis and establishment of the metameric pattern. *Dev. Biol.* **73**, 25-45.

Meier, S. (1980). Development of the chick embryo mesoblast, pronephros, lateral plate and early vasculature. J. *Embryol. exp. Morphol.* **55**, 291-306.

Meier, S. and Tam, P.P.L (1982). Metameric pattern of development in the embryonic axis of the mouse. I. Differentiation of the cranial nerves. *Differentiation.* **21**, 109-122.

Meijlink, F., deLaaf, R., Verrijzer, P., Destree, O., Kroezen, V., Hilken, J. and Deschamps, J. (1987). A mouse homeobox containing gene on chromosome 11: sequence and tissue specific expression. *Nuc. Acid Res.* **15**, 6773-6786.

Melton, D.A. (1987). Translocation of a localised maternal mRNA to the vegetal pole of *Xenopus* oocytes. *Nature.* **328**, 80-82.

Messing, J. (1983). New M13 vectors for cloning. *Methods Enzymol.* **101**, 20-78.

Michiels, F., Burmeister, M. and Lehrach, H. (1987). Derivation of clones close to *met* by preparative field inversion gel electrophoresis. *Science.* **236**, 1305-1308.

Milaire, J. (1971). Prodromes de la genese du membre posterieur chez l'embryon de souris, etude morpholgique et histochimique. *Arch. Biol.* (Bruxelles). **85**.

Mitrani, E., Ziv, T., Thomsen, G., Shimoni, Y., Melton, D.A. and Bril, A. (1990). Activin can induce the formation of axial structures and is expressed in the hypoblast of the chick. *Cell.* **63**, 495-501.

Morata, G. and Lawrence, P. (1977). Homeotic genes, compartments and cell determination in *Drosophila. Nature.* **265**, 211-216.

Nadeau, J.H. (1989). Genome duplication and comparative gene mapping. In *Advanced Techniques in Chromosome Research* (ed. K. Adolph). Dekker, New York.

Nakamura, O., Takasaki, H. and Mizohata, T. (1970). Differentiation during cleavage in *Xenopus laevis*. I. Aquisition of self-differentiation capacity of the dorsal marginal zone. *Proc. Japan Acad.* **46**, 694-699.

Nieuwkoop, P.D. (1969). The formation of mesoderm in Urodele amphibians. I. Induction by the endoderm. *Wilhelm Roux' Arch. EntwMech. Org.* **162**, 341-373.

Nieuwkoop, P.D., Johnen, A.G. and Albers, B. (1985). Inductive interaction and competence. In *The Epigenetic Nature of Early Chordate development* (Developmental and Cell Biology Series: no. 16). Cambridge University press, Cambridge.

Nusslein-Volhard, C., Frohnhofer, H.G. and Lehmann, R. (1987). Determination of anterio-posterior polarity in *Drosophila. Science.* **238**, 1675-1681.

Nusslein-Volhard, C. and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature*. **287**, 795-801.

Otting, G., Qian, Y.Q., Billeter, M., Muller, M., Affolter, M., Gehring, W.J. and Wuthrich, K. (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO. J.* **9**, 3085-3092.

Otting, G., Qian, Y.Q., Muller, M., Affolter, M., Gehring, W.J. and Wuthrich, K. (1988). Secondary structure determination for the *Antennapedia* homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif. *EMBO. J.* **7**, 4305-4309.

Pabo, C.O. and Sauer, R.T. (1984). Protein-DNA recognition. *Ann. Rev. Biochem.* **53**, 293-321.

Percival-Smith, A., Muller, M., Affolter, M. and Gehring, W.J. (1990). The interaction with DNA of wild type and mutant *fushi-tarazu* homeodomains. *EMBO. J.* **9**, 3967-3974.

Pinot, M. (1970). Relations entre le mesenchyme somitique et la plaque des membres chez le poulet. *Annee Biologique*. **9**, 277-284.

Poole, S.J., Kauvaur, L.M., Drees, B. and Kornberg, T. (1985). The *engrailed* locus of *Drosophila*: Structural analysis of an embryonic transcript. *Cell*. **40**, 37-43.

Poustka, A. and Lehrach, H. (1986). Jumping libraries and linking libraries: the next generation of molecular tools in mammalian genetics. *TIG*. **2**, 174-179.

Pratt, H.P.M., Chakraborty, J. and Surani, M.A.H. (1981). Molecular and morphological differentiation of the mouse blastocyst after manipulations of compaction with cytochalasin D. *Cell*. **26**, 279-292.

Ptashne, M. (1988). How eukaryotic transcriptional activators work. *Nature*. **335**, 683-689.

Qian, Y.Q., Billeter, M., Otting, G., Muller, M., Gehring, W.J. and Wuthrich, K. (1989). The structure of the *antennapedia* homeodomain determined by NMR

spectroscopy in solution: comparision with prokaryotic repressors. *Cell.* **59**, 573-580.

Rackwitz, H.R., Zehetner, G., Frischauf, A-M. and Lehrach, H. (1984). Rapid retsriction mapping of DNA cloned in phage vectors. *Gene.* **30**, 195-200.

Rackwitz, H.R., Zehetner, G., Miriald, H., Delius, H., Chai, J.H., Poustka,A., Frischauf, A-M. and Lehrach, H. (1985). Analysis of cosmids using linearisation by phage lambda terminase. *Gene.* **40**, 259-266.

Rawles, M.E. (1963). Tissue interactions in scale and feather development as studied in dermal-epidermal recombination. *J Embryol. exp. Morphol.* **11**, 765-789.

Razin, A. and Riggs, A.D. (1980). DNA methylation and gene function. *Science.* 210, 604-610.

Rebagliati, M.R., Weeks, D.L., Harvey, R.P. and Melton, D.A. (1985). Identification and cloning of localised maternal RNAs from *Xenopus* eggs. *Cell.* **42**, 769-777.

Regulski, M., Harding, K., Kostriken, R., Karch, F., Levine, M. and McGinnis, W. (1985). Homeobox genes of the *Antennapedia* and *bithorax* complexes of *Drosophila. Cell.* **43**, 71-80.

Regulski, M., McGinnis, N., Chadwick, R. and McGinnis, W. (1987). Developmental and molecular analysis of *Deformed*: a homeotic gene controlling *Drosophila* head development. *EMBO. J.* **6**, 767-777

Renard, J-P., Babinet, C. and Barra, J. (1990). Participation of paternal genome is not required before the eight cell stage for full term development of mouse embryos. *Dev. Biol.* **143**, 199-202.

Roderick,T. H. and Hillyard, A. L.(1989). Differences in recombination due to sex in mice. *Mouse News Letter.* **85**, 87.

Rodriguez, R.L. and Tait, R.C. (1983). In *Recombinant DNA Techniques: An Introduction.* Addison-Wesley Publishing Company, Reading, Massachusetts.

Rooze, M.A. (1977). The effects of the *Dh* gene on limb morphogenesis in the mouse. *Birth Defects: Original Article Series.* **13**, 69-95.

Rosa, F.M. (1989). *Mix. 1*, a homeobox mRNA inducible by mesoderm inducers is expressed mostly in presumptive endodermal cells of *Xenopus* embryos. *Cell.* **57**, 965-974.

Rosa, F., Roberts, A.B., Danielpour, D., Dart, L.L., Sporn, M.B. and Dawid, I.B. (1988). Mesoderm induction in amphibians: The role of TGF-β2 like factors. *Science.* **239**, 783-785.

Rosenberg, U.B., Schroder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I. and Jackle, H. (1986). Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIA. *Nature.* **319**,336-339.

Rossant, J. and Frels, W.L. (1981). The origin of the trophoblast and it's role in implantation. In *Cellular and Molecular Aspects of Implantation* (eds. S.R. Glasser and D.W. Bullock). Plenum Press, New York.

Roth, D.B., Porter, T.N. and Wilson, J.H. (1985). Mechanisms of nonhomologous recombination in mammalian cells. *Molec. cell. Biol.* **5**, 2599-2607.

Ruiz i Altaba, A. and Melton, D.A. (1989a). Bimodal and graded expression of the *Xenopus* homeobox gene *Xhox3* during embryonic development. *Development.* **106**, 173-183.

Ruiz i Altaba, A. and Melton, D.A. (1989b). Interaction between peptide growth factors and homeobox genes in the establishment of antero-posterior polarity in frog embryos. *Nature.* **341**, 33-38.

Ruiz i Altaba, A. and Melton, D.A. (1989c). Involvement of the *Xenopus* homeobox gene *Xhox3* in pattern formation along the anterior-posterior axis. *Cell.* **57**,317-326.

Ruiz i Altaba, A. and Melton, D.A. (1990). Axial patterning and the establishment of polarity in the frog embryo. *TIG.* **6**, 57-64.

Russell, E.S. (1985). A history of mouse genetics. *Ann. Rev. Genetics.* **19**, 1-28.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science.* **239**, 487-491.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* (2nd ed.). Cold Spring Harbour Press.

Saxen, L. (1977). Directive versus submissive induction: a working hypothesis. In *Cell and Tissue Interactions* (eds. J. Lash & M. Burger.). Raven Press, New York.

Scheidereit, C., Cromlish, J.A., Gerster, T., Kawakami, K., Balmaceda, C-G., Currie, R.A. and Roeder, R.G. (1988). A human lymphoid specific transcription factor that activates immunoglobulin genes is a homeobox protein. *Nature*. **336**, 551-557.

Schupbach, T. and Wieschaus, E. (1986). Maternal effect mutations affecting the segmental pattern of *Drosophila*. *Wilhelm Roux' Arch. devl. Biol.* **195**, 302-307.

Schwartz, M.N., Trautner, T.A. and Kornberg, A. (1962). Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbour base sequences in deoxyribonucleic acids. *J. Biol. Chem.* **237**, 1961-1967.

Scott, M.P. and Carrol, S.B. (1987). The segmentation and homeotic gene network in early D*rosophila* development. *Cell.* **51**, 689-698.

Scott, M.P. and O'Farrell, P.H. (1986). Spatial programming of gene expression in early *Drosophila* embryogenesis. *Ann. Rev. Cell Biol.* **2**, 49-80.

Scott, M.P. & Wiener, A.J. (1984). Structural relationships among genes that control development: sequence homology between *Antennapedia, Ultrabithorax,* and *fushi-tarazu* loci of *Drosophila*. *Proc. Natl. Acad. Sci. USA.* **81**,4115-4119.

Searle, A.G. (1959). Hereditary absence of the spleen in the mouse. *Nature*. **184**, 1419.

Searle, A.G. (1964). The genetics and morphology of two luxoid mutants in the house mouse. *Genet. Res.* **5**, 171-197.

Shepherd, J.C.W., McGinnis, W., Carrasco, A.E., DeRobertis, E. and Gehring, W. (1984). Fly and frog homeo-domains show homology with yeast mating type regulatory proteins. *Nature*. **310**, 70-71.

Shih, C-C., Stoye, J.P.& Coffin, J.M. (1988). Highly preferred targets for retrovirus integration. *Cell*. **53**, 531-537.

Slack, J.M.W., Dale, L. and Smith, J.C. (1984). Analysis of embryonic induction using cell lineage markers. *Phil. Trans. Roy. Soc. Lond. B.* **307**, 541-549.

Slack, J.M.W., Darlington, B.G., Gillespie, L.L., Godsave, S.F., Isaacs, H.V. and Paterno, G.D. (1989). The role of fibroblast growth factor in early *Xenopus* development. *Development* (Suppl). **107,** 141-148.

Slack, J.M.W., Darlington, B.G., Heath, J.K. and Godsave, S.F. (1987). Mesoderm induction in early *Xenopus* embryos by heparin-binding growth factors. *Nature*. **326**, 197-200.

Slack, J.M.W. and Forman, D. (1980). An interaction between dorsal and ventral regions of the marginal zones in early amphibian embryos. *J. Embrol. exp. Morph.* **56**, 283-299.

Smith, J.C. (1987). A mesoderm inducing factor is produced by a *Xenopus* cell line. *Development*. **99**, 3-14.

Smith, J.C. (1988). Cellular interactions in establishment of regional patterns of cell fate during development. In *Developmental Biology: A Comprehensive Synthesis,* Vol. 5, (ed. L. Browder). Plenum Press, New York.

Smith, J.C., Dale, L. and Slack, J.M.W. (1985). Cell lineage labels and region-specific markers in the analysis of inductive interactions. *J. Embryol. exp. Morphol.* (Suppl). **89**, 317-331.

Smith, C.L., Econome, J.G., Schutt, A., Klco, S. and Cantor, C.R. (1987). A physical map of the *Escherichia coli* K12 Genome. *Science*. **236,** 1448-1453.

Smith, C.L., Lawrence, S.K., Gilliespie, G.A., Cantor, C.R., Weissman, S.M. and Collins, F.S. (1987). Stategies for mapping and cloning macro-regions of mammalian genomes. In *Methods in Enzymology*, Vol. 151, (ed. M.M. Gottesman). Academic Press, San Diego.

Smith, J.C., Price, B.M.J., Van Nimmen, K. and Huylebroeck, D. (1990). Identification of a potent *Xenopus* mesoderm inducing factor as a homolog of Activin A. *Nature*. **345**, 729-731.

Smith, J.C. and Slack, J.M.W. (1983). Dorsalisation and neural induction: properties of the organiser in *Xenopus laevis*. *J. Embryol. exp. Morphol.* **78**, 299-317.

Smith, J.C., Yaqoob, M. and Symes, K. (1988). Purification, partial characterisation and biological effects of the XTC-mesoderm inducing factor. *Development.* **103**, 591-600.

Snell, G.D. and Stevens, L.C. (1975). Early embyology. In *Biology of the Laboratory Mouse* (ed. E.L. Green). Dover Publications Inc., New York.

Sokol, S., Wong, G. and Melton, D.A. (1990). A mouse macrophage factor induces head structures and organises a body axis in *Xenopus*. *Science.* **249**, 561-564.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.

Southern, E.M., Anand, R., Brown, W.R.A. and Fletcher, D.S. (1987). A model for the separation of large DNA molecules by crossed field gel electrophoresis. *Nuc. Acid Res.* **15,** 5925-5943.

Spanopoulou, E., Giguere, V. and Grosveld, F. (1988). Transcriptional unit of the murine *Thy-1* gene: different distribution of transcription initiation sites in the brain. *Mol. Cell Biol.* **8**, 3847-3856.

Spemann, H. (1901). Uber korrelationen in der entwicklung des Auges. *Verh. Anat. Ges., 15 Vers. Bonn* (Anat. Anz.) 15, 61-79.

Spemann, H. & Mangold, H. (1924). Uber induktion von embryonenlagen durch implantation artfremder organisatoren. *Wilhelm Roux' Arch. Entwmech. Org.* **100**, 599-638.

Struhl, G., Struhl, K. and MacDonald, P.M. (1989). The gradient morphogen *bicoid* is a concentration dependent transcriptional activator. *Cell.* **57**, 1259-1273.

Sudarwati, H. and Nieuwkoop, P.D. (1971). Mesoderm formation in the anuran *Xenopus laevis*. *Roux Arch.* **166**, 189-204.

Suzuki, A.S., Yoshimura, Y. and Yano, Y. (1986). Neural-inducing activity of newly mesodermalized ectoderm. *Roux's Arch. Dev. Biol.* **195**, 169-172.

Sweet, H.O. (1982). *Mouse News Letter*. **66,** 66.

Symes, K., Yaqoob, M. and Smith, J.C. (1988). Mesoderm induction in *Xenopus laevis*: responding cells must be in contact for mesoderm formation but suppresion of epidermal differentiation can occur in single cells. *Development*. **104**, 609-618.

Tam, P.P.L. (1981). The control of somitogenesis in mouse embryos. *J. Embryol. exp. Morphol.* **65**, 103-128.

Tam, P.P.L. (1986). A study on the pattern of prospective somites in the pre-somitic mesoderm of mouse embryos. *J. Embryol. exp. Morphol.* **92**, 269-283.

Tam, P.P.L and Beddington, R.S.P. (1986). The metameric organisation of the presomitic mesoderm and somite specification in the mouse embryo. In *Somites in Developing Embryos,* NATO ASI Series, vol. 118.

Tam, P.P.L. and Beddington, R.S.P. (1987). The formation of mesodermal tissues in the mouse embryo during gastrulation and early organogenesis. *Development*. **99**, 109-126.

Tam, P.P.L. and Meier, S. (1982). The establishment of a somitomeric pattern in the mesoderm of the gastrulating mouse embryo. *Am. J. Anat.* **164,** 209-225.

Tam, P.P.L., Meier, S. and Jacobsen, A.G. (1982). Differentiation of the metameric pattern in the embryonic axis of the mouse. II. Somitomeric organisation of the presomitic mesoderm. *Differentiation*. **21**, 109-122.

Tartof, K.D. and Hobbs, C.A. (1987). Improved media for growing plasmid and cosmid clones. *Focus* (Bethesda Research Laboratories). **9**, 12.

Tautz, D. (1988). Regulation of the *Drosophila* segmentation gene *hunchback* by two maternal morphogenetic centres. *Nature*. **332**, 281-284.

Tautz, D., Lehmann, R., Schnurch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. & Jacke, H. (1987). Finger protein of novel structure encoded by *hunchback*, a second member of the gap class of *Drosophila* segmentation genes. *Nature*. **327**, 383-389.

Theiler, K. (1972). *The House Mouse*. Springer Verlag, New York.

Thomas, K.R. and Capecchi, M.R. (1987). Site directed mutagenesis by gene targeting in mouse embryo derived stem cells. *Cell.* **51**, 503-512.

Thomas, K.R., Folger, K.R. and Capecchi, M.R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. *Cell.* **44**, 419-428.

Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L. and Melton, D.W. (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell.* **56**, 313-321.

Thomsen, G., Woolf, T., Whitman, M., Sokol, S., Vaughan, J., Vale, W. and Melton, D.A. (1990). Activins are expressed early in *Xenopus* embryogenesis and can induce axial mesoderm and anterior structures. *Cell.* **63**, 485-493.

Treisman, J., Gonzy, P., Vashishtha, M., Harris, E. and Desplan, C. (1989). A single amino acid can determine DNA-binding specificity of homeodomain proteins. *Cell.* **59**, 553-562.

Tuckett, F., Lim, L. and Morriss-Kay, G.M. (1985). The ontogenesis of cranial neuromeres in the rat embryo. I. A scanning electron microscope and kinetic study. *J. Embryol. exp. Morphol.* **87**, 215-228.

Tykocinski, M. and Max, E.E. (1984). CG nucleotide clusters in MHC genes and in 5' demethylated genes. *Nuc. Acid Res.* **12**, 4385-4396.

van den Eijnden-Van Raaij, A.J.M., van Zoelent, E.J.J., van Nimmen, K., Koster, C.H., Snoek, G.T., Durston, A.J. and Huylebroeck, D. (1990). Activin-like factor from a *Xenopus laevis* cell line responsible for mesoderm induction. *Nature.* **345**, 732-734.

Verbout, A.J. (1985). The development of the vertebral column. *Adv. Anat. Embryol. cell Biol.* **90.**

Viera, J. and Messing, J. (1982). The pUC plasmids an M13amp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene.* **19**, 259-269.

Vincent, J-P. and Gerhart, J.C. (1987). Subcortical rotation in *Xenopus* eggs: An early step in embryonic axis specification. *Dev. Biol.* **123**, 526-539.

Vincent, J-P., Oster, G.F. and Gerhart, J.C. (1986). Kinematics of gray crescent formation in *Xenopus* eggs: The displacement of subcortical cytoplasm relative to the egg surface. *Dev. Biol.* **113**, 484-500.

Warner, A.E. and Gurdon, J.B. (1987). Functional gap junctions are not required for muscle gene activation by induction in Xenopus embryos. *J. Cell Biol.* **104**, 557-564. C

Weeks, D.L. and Melton, D.A. (1987). A maternal mRNA localised to the vegetal hemisphere in *Xenopus* eggs codes for a growth factor related to TGF-β. *Cell.* **51**, 861-867.

Weinzerl, R., Axton, J.M., Ghysen, A. and Akam, M. (1987). *Ultrabithorax* mutation in constant and variable regions of the protein coding sequence. *Genes Dev.* **1**, 386-397.

Weir, M. and Kornberg, T. (1985). Patterns of *engrailed* and *fushi-tarazu* transcripts reveal novel intermediate stages of *Drosophila* segmentation. *Nature.* **318**, 433-439.

Welles, W.L. and Battisto, J.R. (1976). Splenic input on immune capacity of lymphoid cells. In *Immuno-aspects of the Spleen* (eds. J.R. Battisto and J.W. Streiten). Elsevier Press Inc., Netherlands.

Welles, W.L. and Battisto, J.R. (1981). The significance of hereditary asplenia for immunologic competence. In *Immunologic Defects in Laboratory Animals* vol. 1 (eds. M.E. Gershwin and B. Merchant). Plenum Press, New York.

White, R.A.H. and Wilcox, M. (1984). Protein products of the bithorax complex in Drosophila. *Cell.* **39**, 163-171.

Whitman, M. and Melton, D.A. (1989). Growth factors in early embryogenesis. *Ann. Rev. Cell Biol.* **5**, 93-117.

Wilkins, A. (1986). *Genetic Analysis of Animal Development.* John Wiley & Sons, Inc., New York.

Winslow, G.M., Hayashi, S., Krasnow, M., Hogness, D.S. and Scott, M.P. (1989). Transcriptional activation by the *antennapedia* and *fushi-tarazu* proteins in cultured *Drosophila* cells. *Cell.* **57**, 1017-1030.

Wyman, A.R., Wolfe, L.B. & Botstein, D. (1985). Propagation of some human DNA sequences in bacteriophage λ vectors requires mutant *Escherichia coli* hosts. *Proc. Natl. Acad. USA.* **82**, 2880-2884.

Abbreviations

AP	anterior-posterior
ASMP	anterior mesodermal splanchnic plate
bp	base pairs
cDNA	complementary deoxyribonucleic acid
CNS	central nervous system
d	day
ddH_2O	double distilled de-ionised water
dNTP	deoxy nucleoside triphosphate
DNA	deoxyribonucleic acid
ds	double strand
DV	dorso-ventral
dNTP	deoxy nucleoside triphosphate
ds	double strand
Etoh	ethanol
EDTA	ethylenediaminetetra-acetic acid
FGF	fibroblast growth factor
HSV-*tk*	Herpes simplex virus thymidine kinase
hr	hour
ICM	inner cell mass
kb	kilobase pair
Mb	Megabase pair
MIF	mesoderm inducing factor
mRNA	messenger RNA
µl	microlitre
ml	millilitre
µg	microgram
mg	milligram

min	minute
ng	nanogram
^{o}C	degrees celsius
Oligo	oligonucleotide
p.c.	post-coitum
PCR	polymerase chain reaction
PFGE	pulsed field gel electrophoresis
rpm	revolutions per minute
ss	single strand
SDS	sodium dodecyl sulphate
TGF-β	transforming growth factor-β
YAC	yeast artificial chromosome
^{32}p	phosphate isotope, molecular weight of 32