Regulation of Mouse Hoxb-4 Expression During Embryogenesis

Jonathan David Gilthorpe

A thesis presented for the
Degree of Doctor of Philosophy (Ph.D.)
Department of Anatomy and Developmental Biology
University College London

May 1996

National Institute for Medical Research,
Mill Hill, London.
Dedication

For Mum and Dad for setting the course,
for Zöe for trimming the sail,
but most of all for Me for manning the helm.
Abstract

Previous work has defined the sequences sufficient to recapitulate the full expression pattern of the endogenous Hoxb-4 gene in transgenic mice. Several distinct regulatory regions have been identified which are responsible for particular aspects of Hoxb-4 expression. In this study I have begun a detailed analysis of the spatially-specific enhancer, region C, able to drive a major subset of Hoxb-4 expression in the mesoderm, central nervous system (CNS) and peripheral nervous system (PNS). Moreover, it is capable of imposing the correct anterior boundary of Hoxb-4 somitic expression on both the Hoxb-4 and hsp68 promoters. By a combination of sequence comparison and mutational analysis of a region C/hsp68-lacZ reporter gene, I have characterised two cis-regulatory elements that are critical for normal region C activity in transgenic mice. The first of these elements is located within an evolutionarily conserved region of the Hoxb-4 intron (CB1). Deletion of CB1 abolished lacZ reporter gene expression in the mesoderm, PNS and in the majority of the CNS of transgenic embryos. DNA electrophoretic-mobility shift assays revealed the presence of two overlapping binding sites for HoxTF and YY1 within CB1. Specific mutation of each binding-site showed that HoxTF is essential for efficient expression of Hoxb-4 in the embryo, whilst the role of YY1 is unclear. UV crosslinking studies suggest that HoxTF binds to DNA as a heterodimer. Reporter gene analysis has shown that an isolated HoxTF binding element is capable of driving a consistent pattern of expression within the nervous system. These results show that the HoxTF element, though necessary, is not sufficient to drive Hoxb-4 mesodermal expression and requires the cooperation of other elements to achieve this. I have identified a second regulatory element, G5, required to achieve proper levels of region C enhancer activity. A transgenic reporter gene carrying an isolated fragment that contains the HoxTF/YY1 and G5 binding-sites drives a similar pattern of lacZ expression to that seen with the HoxTF site alone. This suggests that further elements are required to specify the Hoxb-4 pattern of mesodermal expression. 3' deletions have mapped these elements to a 269bp fragment that lies 3' to the HoxTF/YY1 site and within the 5' half of the intron. The 3' half of the intron is able to specify a limited pattern of expression in the PNS, implicating the role of a second HoxTF site that is involved in stabilising the activity of the region C enhancer.
Acknowledgements

This thesis constitutes the results of research that was conducted at the National Institute for Medical Research and supported by the Medical Research Council. I would like to thank my supervisor, Dr. Peter W. J. Rigby, for giving me the chance to prove myself and for his inimitable deliverance of advice, guidance and help throughout the course of this work. I am most grateful to Dr. Robb Krumlauf for enthusiastic and helpful discussion. Especial thanks go to Zoë Webster, Lorraine Britton and Hannah Boyes for their champion efforts in the 'mouse house'. I offer my profoundest gratitude to all members of the Rigby and Krumlauf laboratories and friends at the NIMR, colleagues past and present, for advice, support and friendship over the past few years: in particular to Alejandro Gutman, Ed Stanley, Dennis Summerbell, Rajeev Gupta, Oliver Coutelle, Ruedi Allemann, Sui-Pok Yee, Alistair Morrison, Mark Maconochie, Heather Marshall, Kieran Pemberton, Alex Gould, Stefan Nonchev and Linda McNaughton; to Wendy Hatton, John Asante and Paula Timmons for histological assistance; to Jashu Mistry and Alan Harris for the endless supply of oligo's; to the members of the Photographies Division; and to Rod King who is a supreme Director of Studies. A final word of thanks and appreciation to Zoë, whose support, encouragement and understanding has been invaluable to me. I have enjoyed my stay at Mill Hill immensely and will cherish memories, experiences and friendships made here for a long tome to come.
# Table of Contents

Dedication 2  
Abstract 3  
Acknowledgements 4  
Table of Contents 5  
List of Figures 8  
Abbreviations 9  

## Chapter 1: Introduction

1. The molecular basis of pattern formation 11  
   1.1 Pattern formation in *Drosophila* 11  
   1.1.1 Pattern formation in *Drosophila* 11  
   1.1.2 Pattern formation in mouse 13  
1.2 The Homeotic genes 14  
   1.2.1 Homeotic mutations in *Drosophila* 15  
   1.2.2 Activation and maintenance of homeotic gene expression 17  
   1.2.3 The vertebrate *Hox* genes 24  
   1.2.4 Colinearity 29  
   1.2.5 Colinearity, clustering and *Hox* gene regulation 32  
   1.2.6 Mutational analysis of *Hox* genes in the mouse 35  
1.3 The *Deformed*-related *Hox* family 39  
   1.3.1 Characteristics of the *Deformed*-related *Hox* genes 40  
1.4 Regulators of *Hox* gene expression 44  
   1.4.1 *Krox-20* 44  
   1.4.2 *Kreisler* 46  
   1.4.3 Retinoic acid 47  
   1.4.4 *Hox* genes and cofactors 49  
   1.4.5 *Caudal/Cdx* 51  
   1.4.6 *Polycomb* and *trithorax* related factors 53  
1.5 Summary 56  
1.6 Regulatory organisation of the *Hoxb-4* gene 58  
1.7 Aim of the project 59  

## Chapter 2: Materials and Methods

2. Standard solutions and reagents 61  
2.2 DNA manipulations 65  
   2.2.1 Digestion and preparation of DNA for subcloning 65  
   2.2.2 Agarose gel electrophoresis and DNA fragment isolation 66
Chapter 3: Characterisation of a Hoxb-4 region C/heat shock promoter reporter construct

3.1 DNA sequence of Hoxb-4 region C

3.2 Expression of a region C-lacZ reporter gene in transgenic mice

3.3 Summary and Discussion

Chapter 4: HoxTF: a critical factor for Hoxb-4 regulation

4.1 Sequence comparison identifies a conserved region within the intron

4.2 Sequences within CB1 are essential for correct region C activity

4.3 Identification of overlapping binding-sites for HoxTF and YY1 in CB1

4.4 The HoxTF/YY1 binding-site is required for efficient transgene expression

4.5 HoxTF is the major trans-activating component in vivo

4.6 The HoxTF binding-site activates transcription in CAT assays

4.7 Activation by HoxTF in vivo

4.8 HoxTF is widely distributed and binds to DNA as a heterodimer

4.9 Summary and Discussion
5.6 *Cis*-regulatory elements required for mesodermal expression of *Hoxb-4* are located within the left half of the intron. 141

5.7 Summary and Discussion 144

Chapter 6: Final Discussion 157
List of Figures

Chapter 1: Introduction
Fig. 1.1: Summary of HOM-C and Hox gene regulation 20
Fig. 1.2: Homeotic gene expression patterns 27
Fig. 1.3: The Branchial Hox Code 31
Fig. 1.4: Comparison of Dfd-family Hox proteins 43

Chapter 3: Characterisation of a Hoxb-4 region C/heat shock promoter reporter construct
Fig. 3.1: Genomic organisation of the mouse Hoxb-4 gene 80
Fig. 3.2: Sequence of the 1.38kb region C fragment 82
Fig. 3.3: Developmental time-course of region C/hsp68-lacZ reporter gene expression 86

Chapter 4: HoxTF: a critical factor for Hoxb-4 regulation
Fig. 4.1: Sequence conservation between the introns of PG-4 Hox genes 91
Fig. 4.2: CB1 sequences are essential for correct region C activity 94
Fig. 4.3: Localisation of overlapping binding-sites within CB1 98
Fig. 4.4: Model of the mouse Hoxb-4 promoter 100
Fig. 4.5: In vitro analysis of the HoxTF/YY1 binding-site in CB1 103
Fig. 4.6: In vitro analysis of mutations specific for HoxTF and YY1 105
Fig. 4.7: In vivo requirement for the HoxTF/YY1 binding-site in CB1 109
Fig. 4.8: HoxTF mediated trans-activation of region C in vivo 113
Fig. 4.9: The HoxTF binding-site activates transcription in CAT assays 117
Fig. 4.10: The HoxTF binding element is sufficient to drive a subset of the Hoxb-4 neural expression pattern 119
Fig. 4.11: HoxTF is widely distributed and binds to DNA as a heterodimer 123

Chapter 5: The G5 element: an additional cis-regulatory element in region C
Fig. 5.1: 5' Deletions identify a second regulatory element within region C 132
Fig. 5.2: Identification of a G-rich binding element in the bp 475-513 region 135
Fig. 5.3: The Granville binding-site is required for region C function but does not synergise with HoxTF 138
Fig. 5.4: Cis-regulatory elements required for somitic Hoxb-4 expression are located within the left half of the intron 143
Fig. 5.5: In vitro analysis of a second HoxTF/YY1 binding-site in region C 146
Fig. 5.6: Regulatory components of the Hoxb-4 region C enhancer 154
Fig. 5.7: DNA-looping model for the region C enhancer 156
Abbreviations

AP  anterior-posterior
ANT-C  Antennapedia complex
ATP  adenosine 5'-triphosphate
bp  base pair
BSA  bovine serum albumin
BX-C  bithorax complex
CAT  chloramphenicol acetyltransferase
cDNA  complementary DNA
CNS  central nervous system
cpm  counts per minute
C-terminus  carboxy terminus
CV  cervical vertebra
dH2O  distilled water
DNA  deoxyribonucleic acid
DNase  deoxyribonuclease
dATP  deoxyadenosine 5'-triphosphate
dCTP  deoxycytidine 5'-triphosphate
dGTP  deoxyguanosine 5'-triphosphate
dNTP  deoxyribonucleoside triphosphate
dTTP  deoxythymidine 5'-triphosphate
ddNTP  dideoxyribonucleoside triphosphate
dpc  days post coitum
DTT  dithiothreitol
DV  dorsal-ventral
EDTA  ethylenediaminetetraacetic acid
EMSA  DNA electrophoretic-mobility shift assay
HEPES  N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid]
HOM-C  homeotic complex
kb  kilo base (base pairs)
KD  kilo dalton
MBq  mega bequerel
TBq  tera bequerel
M  molar
Mb  mega base (base pairs)
μM  micro molar
mM  milli molar
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>µg</td>
<td>micro gram</td>
</tr>
<tr>
<td>µl</td>
<td>micro litre</td>
</tr>
<tr>
<td>mg</td>
<td>milli gram</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-[N-morpholino]propanesulphonic acid</td>
</tr>
<tr>
<td>Mr</td>
<td>molecular weight</td>
</tr>
<tr>
<td>ng</td>
<td>nano gram</td>
</tr>
<tr>
<td>ml</td>
<td>milli litre</td>
</tr>
<tr>
<td>NP-40</td>
<td>Nonidet P-40</td>
</tr>
<tr>
<td>N-terminus</td>
<td>amino terminus</td>
</tr>
<tr>
<td>OAc</td>
<td>acetate</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>OPA</td>
<td>One-Phor-All buffer</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PG</td>
<td>paralogous group</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulphonyl fluoride</td>
</tr>
<tr>
<td>PNS</td>
<td>peripheral nervous system</td>
</tr>
<tr>
<td>PV</td>
<td>prevertebra</td>
</tr>
<tr>
<td>r</td>
<td>rhombomere</td>
</tr>
<tr>
<td>RA</td>
<td>retinoic acid</td>
</tr>
<tr>
<td>RAR</td>
<td>retinoic acid receptor</td>
</tr>
<tr>
<td>RARE</td>
<td>retinoic acid response element</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>RXR</td>
<td>retinoid X receptor</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>s</td>
<td>somite</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SPF</td>
<td>specific pathogen free</td>
</tr>
<tr>
<td>Tris</td>
<td>2-amino-2-(hydroxymethyl)-1,3-propanediol</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-borate-EDTA buffer</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA buffer</td>
</tr>
<tr>
<td>u</td>
<td>units</td>
</tr>
<tr>
<td>UV</td>
<td>ultra violet</td>
</tr>
<tr>
<td>W</td>
<td>watts</td>
</tr>
<tr>
<td>%(w/v)</td>
<td>grams per 100ml</td>
</tr>
<tr>
<td>%(v/v)</td>
<td>ml per 100ml</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Embryonic development represents a highly coordinated system in which events of cellular proliferation, migration and differentiation are harmonised under a strict genetic control, resulting in a reproducible pattern of cellular diversity that is the adult organism. The elucidation of the molecular mechanisms underpinning this process represents the foremost challenge for developmental biology.

1.1 The molecular basis of pattern formation

Orderly development requires the integration of numerous decisions involving both the overall establishment of the body-plan, or pattern formation, and the discrete specification of cell-type within it. This is dependent on the selective employment of a complex genome, so that specific genes are appropriately regulated according to spatial, temporal and ancestral constraints. Therefore, a key factor in our understanding of the developmental process is a comprehensive knowledge of the regulation of transcription during embryogenesis and the mechanisms by which this is achieved. Central to temporal and spatial genetic control in relation to pattern formation lies the notion of positional information. This concept was pioneered by Wolpert (1969), on which he founded a hypothesis to explain how genetic information gives rise to specific spatial patterns of cellular differentiation. In other words, what are the positional cues from which embryonic cells acquire their particular developmental fates as a result of their location within the maturing embryo? Wolpert's theory involved a bipartite mechanism whereby cells within a developing system would be informed of their position with respect to certain reference points, as a result of positional information. Having acquired a given positional value, the cells would then interpret this information according to their lineage by adopting a particular developmental pathway (see Summerbell et al., 1991). Much of our current understanding of the biological processes involved in pattern formation has been derived from the study of Drosophila melanogaster.

1.1.1 Pattern formation in Drosophila

The development of the fertilised egg incorporates several analogous features in both Drosophila and mouse embryos, although a number of important differences exist. Most
notable is the patterning role of maternally derived factors present in the *Drosophila* egg prior to fertilisation.

The *Drosophila* egg is asymmetrically shaped with anterior and posterior poles and dorsal and ventral surfaces that are clearly distinct. Four maternal organiser systems (anterior, posterior, terminal and dorsoventral) are required to establish the basic body pattern along the anterior-posterior (AP) and dorsal-ventral (DV) embryonic axes. Each system provides the activity of a spatially distributed determinant in the form of a transcription factor (see Hoch & Jäckle, 1993; Kornberg & Tabata, 1993). For example, in the anterior system maternally deposited *bicoid* (*bcd*) mRNA is sharply localised at the anterior tip of the egg (Belerth *et al*., 1988). After fertilisation the *bcd* protein diffuses away from the anterior pole to form a roughly exponential AP concentration gradient (Driever & Nüsslein-Volhard, 1988). The *bcd* protein elicits distinct transcriptional responses from subordinate regulatory genes towards the anterior end of the embryo, according to different relative values of the *bcd* protein gradient Hoch & Jäckle, 1993; Jäckle & Sauer, 1993. In this way *bcd* is thought to function in the specification of anterior positional information.

At the zygotic stage *Drosophila* embryogenesis proceeds *via* a state of superficial cleavage, during which a rapid succession of nuclear divisions occurs without cellularisation. The nuclei then migrate out to the periphery of the embryo forming a syncytial layer (the syncytial blastoderm) and, after a further period of division, the nuclei become surrounded by membranes to form the cellular blastoderm (see Lawrence, 1992). This period of acellularity permits the free diffusion of factors provided by the maternal systems. Just prior to cellularisation a cascade of regulatory genes is activated in response to the maternally derived stimuli. In relation to the formation of the AP axis, all of the early genes of this cascade encode transcription factors, the target genes of which include other members of the hierarchy. According to their phenotypic and temporal effects these so called segmentation genes can be grouped into three classes, i.e. the gap, pair-rule and segment-polarity genes. One of the effects of this cascade is to translate the relatively broad domains of gap gene expression along the AP axis of the syncytial blastoderm into a more refined, striped pattern of pair-rule and segment-polarity gene expression in the cellular blastoderm. Products of the pair-rule and segment-polarity genes act in a cooperative manner to initially define the metameric pattern of the embryo (see Ingham, 1988; Nüsslein-Volhard, 1991; Lawrence, 1992; Hoch & Jäckle, 1993; Jäckle & Sauer, 1993; Kornberg & Tabata, 1993).

Following gastrulation, during which cells along the dorsal midline invaginate to form the mesodermal layer, segmentation of the embryo occurs by a series of condensations of the mesoderm and corresponding indentations of the ectoderm. This process establishes the typical larval body pattern consisting of a number of homologous metameric units termed
parasegments. The parasegments predispose the formation of the characteristic head, thoracic and abdominal segments of the adult fly. The combined activities of the gap and pair-rule genes, which act in the syncytial phase of *Drosophila* embryogenesis, initiate the expression of both segment-polarity and homeotic genes to bring about the distinctive segmental pattern seen later in the larva (see Pankratz & Jäckle, 1993). The sequential activation of each class of gene indicates that, in *Drosophila* at least, a hierarchical network of gene expression and cell commitment generates the final body pattern.

### 1.1.2 Pattern formation in mouse

Some parallels may be derived between mammalian and *Drosophila* embryogenesis. The fertilised egg undergoes an initial cleavage period followed by the formation of an ectodermal layer. Further analogies may be drawn with the formation of the three germ layers that make up the embryo; ectoderm, mesoderm and endoderm. To a lesser extent vertebrates also show evidence of metameric organisation, particularly in the somites of the paraxial mesoderm and in the nervous system (see Hogan *et al.*, 1994). Despite these conformities several notable differences exist. As previously mentioned, in the unfertilised *Drosophila* egg there is an initial state of polarisation due to maternal components. In the mouse there is little or no obvious polarisation of the embryo up until the formation of the primitive streak at the onset of gastrulation. It is thought that at around this stage the fundamental decisions as to the definition of the AP and DV axes are made, though the mechanisms involved in these processes are unclear. Another noteworthy contrast is that in *Drosophila* nuclear division occurs during an acellular state, allowing the diffusion of factors throughout the early blastoderm. The cellular system of the mouse embryo must therefore require more complex mechanisms, involving inductive interactions and cell-cell communication, for the generation and dissemination of positional information.

The absence of suitable genetic screens has hindered study of the molecular mechanisms underlying key processes in murine development, such as pattern formation. Mutations in prominent developmental control genes, acting early on in mammalian embryogenesis, are likely to produce embryonic-lethal phenotypes which can not be easily scrutinised. Many developmental mutations have been generated in *Drosophila*, resulting in the identification of numerous genes that dictate the events of development and their associated pathways. In the hope of distinguishing similar factors that operate in vertebrate species, the realisation that many important determinants of *Drosophila* development share conserved domains has allowed us to screen for homologous counterparts. A large number of genes and gene families involved in regulating fundamental mechanisms during development have been identified in this way from many diverse species, including those encoding; transcription
Comparative biology and evolutionary analysis have grown in stature as important tools for dissecting the underlying genetic themes governing embryogenesis. Despite these efforts it is uncertain to what extent the functional conservation of evolutionarily related genes has been preserved. What is clear, however, is that whilst structural similarities may indicate properties that are required for a biochemical function within a given cellular or developmental process (e.g. DNA binding domains), this does not necessarily extrapolate to a conserved role in development per se.

1.2 The Homeotic genes

Some of the most extensively characterised regulators of pattern formation are the Homeotic genes of Drosophila. Genes of this class were identified by virtue of their characteristic homeotic mutant phenotypes, where the identity of one segment, or a group of them, is transformed into the likeness of another. Their role is in the establishment of appropriate segmental identities during both embryonic and imaginal development, by encoding determinants that act as molecular switches to define the developmental pathway a particular segment will adopt (Lewis, 1978; Mahaffey & Kaufman, 1987; Morata, 1993).

The term 'homeosis' first entered the limelight in 1894 when William Bateson, in his work Material for the Study of Variation, Treated with Especial Regard to Discontinuity in the Origin of Species, used it to describe the biological variation in which 'something has been changed into the likeness of something else' (Bateson, 1894; see Lewis, 1994). Bateson pointed to the possible significance of homeotic transformations for developmental biology and highlighted that such altered pathways of development are present in many metazoan species. The first established case of a homeotic mutant was described by Bridges & Morgan (1923) in Drosophila. The phenotype of this bithorax (bx) mutation involves a weak conversion of the third thoracic segment into the second, resulting in flies that carry a small second pair of wings in place of halteres. Further homeotic mutants were subsequently identified, many of which map to two independent but linked genetic loci, namely the Antennapedia (ANT-C) and bithorax (BX-C) complexes (Lewis, 1978; Kaufman et al., 1980, 1990), collectively referred to as the Homeotic complex (HOM-C) (Akam, 1987, 1989).

A total of eight homeotic genes constitute the HOM-C: labial, lab; proboscipedia, pb; Deformed, Dfd; Sex combs reduced, Scr; Antennapedia, Antp; Ultrabithorax, Ubx;
abdominal A, abd-A; and Abdominal B, Abd-B. Sequence comparison between the homeotic genes revealed that they share a striking 180bp region of homology (McGinnis et al., 1984a; Scott & Weiner, 1984). This region, termed the homeobox, encodes the 60 amino acid homeodomain motif that binds to DNA in a sequence specific manner and is found in many transcription factors from eukaryotic species (Laughon & Scott, 1984; McGinnis et al., 1984a, 1984b; Scott et al., 1989; Laughon, 1991; Gehring et al., 1994; Mann, 1995; Stein et al., 1996). Attempts have been made to classify homeobox genes according to their relatedness, both within the homeodomain and in other conserved regions external to it. This has allowed the evolutionary relationships that exist between them to be determined and assisted in structural and functional studies (see Bürglin, 1994). The largest group is the Antennapedia class, encompassing homeodomains that share approximately 60% identity with that first identified in the *Drosophila Antp* gene, including all of the HOM-C genes. Members of this class possess an identical core sequence of 12 amino acids between positions 44-55 in α-helix III of the homeodomain (Scott et al., 1989; Gehring et al., 1990). In addition they also usually contain a short conserved pentapeptide or hexapeptide motif containing the core sequence YPWM, encoded by the exon immediately 5' of the homeodomain exon (Mavilio et al., 1986; Krumlauf, 1992).

1.2.1 Homeotic mutations in *Drosophila*

Two types of homeotic mutation have been described which have contrasting effects. Recessive (null, loss-of-function) mutations are due to the inactivation or deletion of the gene in question, resulting in the absence of the corresponding gene product and associated segmental defects. The identity of the affected segment(s) becomes transformed into that of a neighbouring one. Often this is visible in the resulting larval cuticle as duplications of certain pattern elements with a coincident loss of others. Dominant (gain-of-function) mutants arise from the inappropriate, i.e. constitutive or ectopic, localisation of the intact gene product, changing the identity of segments that would not normally express it (Lewis, 1978; Wakimoto & Kaufman, 1981; Hayes et al., 1984; Sato et al., 1985; Gibson & Gehring, 1988; González-Reyes & Morata, 1990).

The first homeotic mutations to be characterised at the molecular level were those of *Antp*. Null *Antp*-mutations result in a transformation of the identity of the second thoracic segment (T2) into that of the first thoracic segment (T1). Since this is an embryonic-lethal phenotype it can only be studied in the adult by inducing mutant clones locally and comparing their cuticular features with those of wild-type areas (Struhl, 1981; Wakimoto & Kaufman, 1981). Conversely, the dominant gain-of-function mutation (hereafter designated as AntpDp) results in the opposite effect. T1 normally generates the antennae of the adult and the phenotype of
Antp\textsuperscript{D} is a replacement of the antennae by a second set of legs, normally specified by T2. Antp activity is needed for normal metameric identity in the thoracic region and presumably it is specifically not required in the anterior head. Inappropriate expression of Antp in this region would therefore result in the acquisition of thoracic characteristics, such as is seen. This interpretation has been directly confirmed by introducing an artificial Antp\textsuperscript{D} mutation, consisting of the Antp cDNA coupled to a heat shock promoter, via a P-element construct into the germline of wild-type flies. When individuals carrying this construct were exposed to ubiquitous Antp activity by heat induction, the anticipated transformations from antennae to legs were observed (Schneuwly \textit{et al.}, 1987a). Research into Antp\textsuperscript{D} has shown that the Antp protein-coding region is inappropriately expressed in the anterior head. This is caused by an inversion mutation in which the protein-coding unit of Antp becomes fused to the 5' exons of another gene (rfd) that is transcribed in the opposite sense. This chimeric rfd/Antp gene is expressed in the head (the normal domain of rfd expression) causing the characteristic antenna to leg transformation (Frischer \textit{et al.}, 1986; Schneuwly \textit{et al.}, 1987b).

The genes of \textit{ANT-C} and \textit{BX-C} govern positional identity in the developing anterior and posterior segments of the embryo, respectively (Lewis 1978; Kaufman \textit{et al.}, 1980, 1990; Harding \textit{et al.}, 1985). The diversity of the segmental pattern depends on the spatial activity of the \textit{HOM-C} genes which act alone or in combination (Lawrence & Morata, 1994). Their gene products can bind to and regulate other genes, which in turn are involved in determining segmental structure and function. The identification of such downstream target genes has proved elusive, however, and efficient \textit{in vivo} techniques for this purpose are somewhat lacking (Morata & Struhl, 1990; Andrew & Scott, 1992). One approach has been to map the sites where \textit{HOM-C} proteins bind to polytene chromosomes. Other strategies, employing specific antibodies for chromatin-immunoprecipitation (Gould \textit{et al.}, 1990; Graba \textit{et al.}, 1992), or the genetic selection of potential target sites from a library of genomic DNA fragments in yeast (Mastick \textit{et al.}, 1995), have been successful in identifying a few genes whose expression patterns fit those predicted. Immunoprecipitation with antibodies directed against Ubx has led to the identification of \textit{connectin} as a target (Gould \textit{et al.}, 1990; Gould & White, 1992). This gene encodes a cell-surface protein that may play a role in the recognition of target muscles by motor neurons (Nose \textit{et al.}, 1992). A more indirect method involves the use of enhancer-trap genetics to identify patterns of expression that are dependent on the activity of a particular homeotic gene (Wagner-Bernholz \textit{et al.}, 1991; Mahaffey \textit{et al.}, 1993). This line of study has lead to the identification of the \textit{spalt major} gene, encoding a protein with zinc finger DNA-binding motifs (Wagner-Bernholz \textit{et al.}, 1991). Indeed, several of the downstream target genes that have been so far identified are either transcriptional regulators, or secreted factors that mediate cell-cell interactions (Andrew & Scott, 1992; Botas, 1993). This may indicate that at least some of the \textit{HOM-C} genes act, in part, by governing a subordinate cascade of secondary transcription factors.
1.2.2 Activation and maintenance of homeotic gene expression

The homeotic genes exhibit complex and dynamic patterns of expression during embryonic development. Diverse mechanisms are thought to be involved in the initial activation phase and the maintenance of the 'on/off' state of expression, mediated by cis-acting elements which can be negatively or positively stimulated by a range of upstream factors.

Initial establishment of homeotic gene expression occurs in spatially restricted domains along the AP axis of the embryo in response to positional information. This is provided by the concentration of proteins encoded by the early acting segmentation genes, such as the zinc-finger protein products of the gap genes hunchback (hb) and krüppel (kr). For example, the domain of Ubx expression is initially activated in a broad stripe in the middle of the embryo, between certain values of the hb gradient (Struhl et al., 1992). Subsequently the limits of homeotic gene expression sharpen and are brought into register with the anterior parasegmental boundaries. This requires the activities of several other transiently expressed segmentation genes such as fushi tarazu (ftz) and even skipped (eve) (Ingham & Martinez-Arias, 1986; Irish et al., 1989; Struhl et al., 1992). These factors have been shown to specifically bind to DNA sequences located within the regulatory regions of several homeotic genes including Ubx (Qian et al., 1991; Zhang et al., 1991; Müller & Bienz, 1992; Shimell et al., 1994).

Equally important for the function of the homeotic genes is the continued maintenance and modulation of their expression in the post-establishment phase. The definition of the normal body-plan requires the continuous expression of each homeotic gene throughout development within the domains defined at earlier stages (Morata & García-Bellido, 1976; Struhl, 1982; Lawrence, 1992). However, expression of the early-acting segmentation genes subsides during gastrulation and additional factors are required to sustain homeotic gene expression patterns during later stages. This is achieved by a combination of positive and negative regulators of homeotic gene expression and by autoregulatory and crossregulatory interactions among homeotic genes themselves (see Figure 1.1).

At least two groups of global trans-acting factors serve to maintain the active or repressed state of the homeotic genes after their early domains of expression have been established. The factors responsible are thought to exert their effects by the alteration of chromatin structure. Genes of the Polycomb group (Pc-G) function to repress transcription and were originally identified due to their homeotic mutant phenotypes (Busturia & Morata, 1988). In Pc-G mutant embryos, homeotic genes are initially expressed within appropriate domains but become ectopically expressed by mid-embryogenesis, suggesting that the Pc-G genes are required to maintain but not establish the domains of homeotic gene expression (Struhl &

Pc-G proteins are thought to repress homeotic gene activity via the formation of large complexes that bind to specific sites within HOM-C and induce the formation of heterochromatin. In support of this theory, recent research has shown that several Pc-G members including Polycomb (Pc), polyhomeotic (ph), Posterior sex combs (Psc) and Polycomb-like (Pcl), encode polytene chromosome associated proteins. This association requires the presence of a conserved protein motif, the chromodomain (Zink & Paro, 1989; DeCamillis et al., 1992; Messmer et al., 1992; Martin & Adler, 1993; Rastelli et al., 1993; Lonie et al., 1994; Chang et al., 1995a). Although the precise mechanism of this process are unclear, the tight packaging of chromatin-associated DNA would prohibit the functional interaction of activating transcription factors with their target sites, thereby silencing homeotic gene expression (reviewed in Paro, 1990, 1993). Recent investigations have indicated that Pc-G-mediated repression may act in a selective manner, allowing promoter-access to certain components of the transcriptional machinery but not to others (McCall & Bender, 1996). Analysis of the genetic interactions between various Ubx and Pc mutations suggests that the effects of Pc are primarily mediated through distal regulatory elements, located many kilobases upstream and downstream of the transcription start site (Castelli-Gair & García-Bellido, 1990).

Attempts so far have failed to show that Pc and ph proteins are able to bind to DNA in a sequence-specific manner (Zink et al., 1991; Franke et al., 1992). A reporter gene driven by the BXD enhancer of Ubx, which directs extensive expression throughout most of embryogenesis, may be restricted to the Ubx domain by the addition of isolated hb binding-sites, or Ubx control regions containing them. This has prompted the hypothesis that Pc-DNA associations are mediated via the tethering effect of protein-protein interactions, with hb representing one of these tethering factors. This silencing effect acts at long range, is dependent on Pc activity, and is maintained after stages in which hb protein is no longer detectable (Müller & Bienz, 1991; Zhang et al., 1991; Zhang & Bienz, 1992; Busturia & Bienz, 1993). More recent evidence has shown that transcriptional silencing can be achieved by tethering Pc to DNA in a sequence-specific manner via the DNA binding domain of the yeast transcription factor GAL4 (i.e. as a GAL-Pc fusion protein). In this case it was demonstrated that silencing of the Ubx BXD enhancer is independent of hb activity and requires the C-terminal region of Pc but not the chromodomain. Synthetic reporter genes containing GAL4 binding-sites could only be transiently repressed by a pulse of GAL-Pc. In combination with the BXD enhancer, however, stable silencing was achieved in a Pc-G gene dependent fashion (Müller, 1995).

In summary, Pc-G genes are required for the stable and heritable silencing of homeotic gene
Figure 1.1: Summary of HOM-C and Hox gene regulation

Functional and structural analyses of the *Drosophila* HOM-C and vertebrate Hox complexes suggests a common role in embryogenesis that has been maintained throughout evolution. It is presumed, therefore, that common pathways may be involved in gene regulation. Hox/HOM genes are thought to act on groups of downstream target genes that govern similar cellular processes during development. Some of these targets and/or processes may be related as indicated by a reversible arrow.

Analogous mechanisms are also thought to be employed in the maintenance phase of Hox/HOM gene expression, such as autoregulation and the involvement of genes belonging to the Polycomb (Pc-G) and trithorax (trx-G) groups. Pc-G and trx-G proteins are believed to act antagonistically to modify chromatin structure, heritably maintaining the appropriate patterns of Hox/HOM gene expression once they have been established. Pc-G factors have a repressive role, possible by inducing the formation of heterochromatin to silence transcription, whilst trx-G factors have an activating role, probably through the preservation of an open-chromatin state.

The initial activation of HOM-C gene expression occurs during a syncytial phase of development whereby the activity of diffusible transcription factors can directly stimulate their expression. In contrast, vertebrate Hox gene activation may depend on upstream factors that are distinct from those of *Drosophila* (indicated by a hashed reversible arrow) and involve intercellular signalling mechanisms.
Drosophila

Vertebrates

Upstream Factors

Pc-G

trx-G

autoregulation

HOM-C

Target Genes

Hox

Target Genes

Pc-G

trx-G

autoregulation

\[ \text{autoregulation} \]
expression in Drosophila. It appears that initial repression governed by the transiently expressed early-acting segmentation genes (e.g. hunchback) becomes "locked in" by a mechanism involving the recruitment of Pc-G proteins to form nucleoprotein complexes via protein-protein interactions. This silencing process apparently requires the anchoring of Pc-G proteins to specific cis-regulatory sequences present within the homeotic genes and may involve multiple functional domains in these proteins, including both the chromodomain and the C-terminal region.

The second second set of global trans-acting factors that are involved in the maintenance phase of homeotic gene expression constitute the trithorax group (trx-G). Genes of the trx-G, such as trithorax itself, encode activators of homeotic gene transcription. These factors play an antagonistic role to that of the Pc-G genes, being required for the maintenance of appropriate levels of expression of each homeotic gene within its normal domain (Capdevila & García-Bellido, 1981; Shearn et al., 1987; Kennison & Tamkun, 1988; Shearn, 1989; Tamkun et al., 1992). Mutations in trx cause wide ranging homeotic transformations suggesting a role in regulating multiple HOM-C genes (Breen & Harte, 1993).

The mechanisms of activation by the products of trx-G genes are unclear but may involve the drawing together of distant cis-regulatory elements and their associated promoters via DNA-looping (reviewed in Kennison, 1993). Several lines of evidence suggest that the products of trx and Pc undergo competitive interactions. It has been observed that embryos and adults of trx/Pc- double mutants exhibit fewer abnormalities that those bearing the Pc mutation alone (Capdevila & García-Bellido, 1981; Sato & Denell, 1987). Many trx alleles were uncovered in a systematic screen to isolate suppressors of Pc (Kennison & Tamkun, 1988). Mutations in trx seem to interact with the same set of regulatory mutations of Ubx as do Pc mutations (Castelli-Gair & García-Bellido, 1990). Activation of a reporter gene linked to an Ubx enhancer fragment can be mediated by trx but completely abolished by Pc in a dose-dependent manner (Chang et al., 1995a). It is therefore possible that competitive interactions between trx-G and Pc-G proteins, and cis-regulatory sequences, determine the active or repressed state of HOM-C gene transcription.

Molecular analyses have shown that trx encodes at least three predicted protein isoforms produced by differential splicing, the largest of which is 3,759 amino acids in length (Mazo et al., 1990; Breen & Harte, 1991). Structurally these proteins contain several clusters of unusual cysteine-rich zinc-binding motifs in the central region, suggesting that trx might bind to DNA. This domain is required for trx-dependent reporter gene activation (Chang et al., 1995a). The brahma (brm) gene, which is another trx-G member, was first identified as a suppressor of Pc mutations. The protein encoded by brm shares extensive homology with a family of DNA-dependent ATPases and helicases, including the yeast SNF2/SWI2 protein
that belongs to a group of global transcriptional activators (Kennison & Tamkun, 1988; Tamkun et al., 1992; Henikoff, 1993). SNF2/SWI2 does not appear to directly bind to DNA, rather it seems to assist DNA-binding regulatory proteins to overcome the repressive effects of chromatin structure on transcription (Peterson & Herskowitz, 1992). Based on the similarities between brm and SNF2/SWI2, it has been suggested that brm assists trx to overcome repressive effects on HOM-C transcription (Tamkun et al., 1992).

The HOM-C genes are also subject to transcriptional regulation by autoregulatory and crossregulatory mechanisms. Sustained repression of each HOM-C gene outside of its normal expression domain is mediated by crossregulatory interactions among the HOM-C genes themselves (Struhl, 1982; Hafen et al., 1984; Harding et al., 1985; Struhl & White, 1985; Carroll et al., 1986). HOM-C gene activity may be maintained by autoregulatory mechanisms (reviewed in Bienz, 1994). Direct autoregulation is mediated via the homeodomain of the protein directly binding to cis-regulatory sequences within the same gene and providing a positive feedback loop (Jiang et al., 1991; Regulski et al., 1991; Schier & Gehring, 1992). Feedback may also occur indirectly, being dependent on intermediary extracellular signalling molecules derived from neighbouring cells (Heemskerk et al., 1991; Tremml & Bienz, 1992; Thuringer et al., 1993).

The first evidence for direct autoregulation in Drosophila came from studies on the homeobox-containing segmentation gene fushi tarazu (ftz) (Schier & Gehring, 1992). In vitro and in vivo experiments demonstrated that the activity of a distal ftz enhancer region, that specifically binds the ftz protein, was inhibited when mutations were introduced into putative ftz binding-sites or into the homeodomain of the ftz protein itself. Compensatory second-site suppressor analysis (Jarvik & Botstein, 1975) was used to demonstrate that ftz autoregulation occurs by a direct mechanism. In brief, this method involves the introduction of a mutation into a putative transcription factor binding site, the phenotype of which can be suppressed by a compensatory mutation within the factor of interest. In this case a motif within the ftz enhancer was mutated so that it was recognisable by the homeodomain of bcd. A compensatory bcd-like mutation was introduced into the ftz homeodomain. This modified ftz protein, now capable of binding to the modified site, was able to suppress the binding-site mutation and stimulate expression. Had ftz autoregulation occurred by an indirect mechanism, activation would not have occurred.

One of the best characterised examples of HOM-C autoregulation involves the Deformed gene, responsible for specifying the mandibular and maxillary segments of the posterior head region in Drosophila (McGinnis et al., 1990). Null mutations in Dfd have an embryonic-lethal phenotype, resulting in the loss of cuticular specialisations of the mandibular and maxillary segments and the duplication of cuticular elements of the larval head skeleton.
(Regulski et al., 1987). The idea that Dfd auto-activation plays an important role in the maintenance of its own expression is drawn from two lines of evidence. Firstly, Dfd expression is activated appropriately in embryos carrying nonsense codons in the Dfd open reading frame. However, normal levels of Dfd transcription are not maintained in either epidermal or neural cells of these embryos (Kuziora & McGinnis, 1988). Secondly, ectopic localisation of the Dfd protein via a heat-inducible expression construct can activate endogenous Dfd transcription in some cells of the thoracic and abdominal epidermis. Ectopic Dfd expression is maintained for much of embryogenesis and induces the formation of ectopic maxillary cuticular structures. The DNA sequences responsible for this epidermal aspect of Dfd autoregulation have been mapped to a large upstream enhancer. A small, segment-specific regulatory element within this enhancer contains binding sites for Dfd and for a cofactor (DEAF-1), both of which are required for autoregulation (Bergson & McGinnis, 1990; Zeng et al., 1994; Gross & McGinnis, 1996). More recently an autoactivated enhancer element situated in one of the Dfd introns has been implicated in the maintenance of Dfd expression within the embryonic central nervous system (Lou et al., 1995).

An example of indirect feedback is demonstrated by Ubx (reviewed in Bienz, 1994). Studies have shown that Ubx expression within the visceral mesoderm of the Drosophila embryo depends not only on the activity of Ubx but also on the activities of the extracellular signalling molecules decapentaplegic (dpp) and wingless (wg) (Bienz & Tremml, 1988; Panganiban et al., 1990; Hursh et al., 1993; Thuringer & Bienz, 1993). Both dpp and wg are expressed within adjacent domains of the visceral mesoderm and their expression is dependent on Ubx activity (van den Heuvel et al., 1989; Panganiban et al., 1990). Other observations suggest that there is also a direct component to the autoregulation of Ubx in this context. Ubx expression is completely abolished in null Ubx mutant embryos. However, in dpp/wg double mutants this expression is merely down-regulated (Immergluck et al., 1990; Panganiban et al., 1990; Thuringer et al., 1993). The purpose of such a double mechanism of autoregulation and feedback remains unclear. It is possible that direct autoregulation alone is a relatively unstable mechanism for achieving synchronous Ubx expression within a given cellular population, being sensitive to minor fluctuations in the levels of local protein concentration. Thus, the incorporation of an indirect feedback pathway might act to stabilise the coordinate expression of Ubx within cells of the visceral mesoderm. It is thought that such non-cell-autonomous mechanisms of transcriptional regulation, with HOM-C genes being able to feedback on their own expression in neighbouring cells, is not a general feature of their activity (see Lawrence & Morata, 1994). If it were then cell/segmental identity specification, which is a salient characteristic of homeotic gene function, could surely not be achieved. This may explain why both the direct and indirect pathways of Ubx autoregulation coexist. Since Ubx activity is absolutely required for autoregulation, this provides a
safeguard by which *Ub* activation can not be triggered in non-*Ub* expressing cells by the indirect mechanism.

The only homeotic gene that is known to be expressed in the endoderm of the *Drosophila* embryo is *labial* (Diederich *et al*., 1989). The expression domain of *lab* within the midgut epithelium occupies a single band directly underlying the domain of *Ub* expression in the visceral mesoderm. The endodermal expression of *lab* is induced by extracellular factors emanating from the adjacent visceral mesoderm and is therefore governed by an indirect feedback mechanism (Tremml & Bienz, 1992). It has been demonstrated that the *lab* expression domain is dependent on *Ub* via the activity of *dpp* and *wg* (Immergluck *et al*., 1990; Panganiban *et al*., 1990; Reuter *et al*., 1990). Identification of 5′ *lab* regulatory sequences revealed the presence of two response elements. The activity of the first is dependent on the *dpp* gene product, whilst the second is *lab*-dependent, indicating that both direct autoregulatory and indirect feedback mechanisms cooperate to maintain *lab* expression in the developing embryo.

Soon after its discovery, the prototypic homeobox from *Antp* was used as a probe to screen for homeobox-containing genes in other species (Carrasco *et al*., 1984; Levine *et al*., 1984; McGinnis *et al*., 1984a, 1984b; Müller *et al*., 1984; Colberg-Poley *et al*., 1985). It has now become evident that there are true homologues of the *HOM-C* genes in vertebrates that share conserved roles in the specification of regional identity along the AP axis during embryogenesis (see McGinnis & Krumlauf, 1992; Krumlauf, 1994).

### 1.2.3 The vertebrate *Hox* genes

The homeodomain has been identified in a wide range of eukaryotic proteins and consequently has become one of the most studied DNA-binding motifs (Laughon, 1991; Duboule, 1994; Gehring *et al*., 1994). The vertebrate *Hox* genes, encoding Antennapedia class homeodomain proteins, have been extensively characterised and are most closely related to the *HOM-C* genes of *Drosophila* by both sequence similarity and their clustered genomic organisation (Gaunt *et al*., 1988; Holland & Hogan, 1988; Boncinelli *et al*., 1989; Duboule & Dollé, 1989; Graham *et al*., 1989; Kessel & Gruss, 1990). It is now widely accepted that the *Hox/HOM* genes represent an intrinsic mechanism for the specification of positional values along the AP body axis during embryogenesis. This dogma has arisen from the characterisation of equivalent systems that are present throughout the metazoan animal subkingdom (McGinnis & Krumlauf, 1992; Kenyon, 1994; Krumlauf, 1994).

Like the process of segmentation in *Drosophila*, a fundamental necessity of vertebrate
development is the generation of serially repeated or metameric units along the AP axis of the embryo. In mouse, the earliest morphological semblance of a segmental pattern is seen with the formation of neuromeres in the neural tube, and the appearance of somites in the paraxial mesoderm (see Hogan et al., 1994). Both of these events occur immediately after the completion of germ layer formation (Tam, 1981; Jacobson & Tam, 1982). Segmentation of the neuroectoderm and the mesoderm has a direct impact on the patterning of the head and trunk. In the neural tube, the neuromeric pattern governs the spatial organisation of neuronal tracts within the brain and the migration of neural crest into the branchial arches (Lumsden 1990; Hunt et al., 1991b; Lumsden et al., 1991; Fraser, 1993). In the trunk, the somitic pattern provides the foundations for the formation of the vertebral column, ribs and the axial musculature (Verbout, 1985; Buckingham, 1992), as well as influencing the architecture of the peripheral nervous system and the migration of neural crest (Stern & Keynes, 1987; Stern et al., 1991). The vertebrate Hox genes are predominantly expressed in the derivatives of the ectoderm, in the central and peripheral nervous system, and in the mesoderm (Fienberg et al., 1987; Krumlauf et al., 1987; Graham et al., 1988), displaying similar overlapping domains of expression within any given tissue (Gaunt, 1988; Graham et al., 1988; Holland & Hogan, 1988). This is most strikingly observed in the developing central nervous system and somites, where the anterior expression domains of certain Hox genes coincide with specific segmental boundaries, e.g. boundaries between hindbrain rhombomeres (Wilkinson et al., 1989b).

Our most detailed knowledge of the vertebrate Hox genes stems from analyses of the mouse and human Hox complexes. This indicates that there are 38 individual Hox genes ordered into four unlinked clustered arrays of 9-11 genes. The constituent genes of each cluster are orientated in the same 5' to 3' sense with respect to the direction of transcription. In the mouse the four clusters are known as Hoxa, Hoxb, Hoxc and Hoxd and are located on chromosomes 6, 11, 15 and 2, respectively. The nomenclature of the human Hox genes follows the same convention but uses capital letters to distinguish them from the murine clusters, hence Hoxa is designated HoxA in humans and so on. Several studies on diverse species indicate that this organisational arrangement of Hox genes is common to all vertebrates. Members of other metazoan phyla such as arthropods and nematodes, as well as lower chordate sub-phyla such as the cephalochordates, seem to have only one Hox complex. Thus, it has been postulated that changes in Hox gene number and genomic organisation played a crucial role in the evolution of the metazoan body-plan (Akam, 1989; Kappen et al., 1989; Kessel & Gruss, 1990; Duboule, 1992; Holland, 1992; Krumlauf, 1992; McGinnis & Krumlauf, 1992; Scott, 1992; Duboule, 1994; Holland & Garcia-Fernández, 1996).

The striking organisational and homology relationships between the vertebrate Hox clusters and HOM-C of Drosophila are illustrated in Figure 1.2. These affiliations are based on
Figure 1.2: Homeotic gene expression patterns

The organisational and functional similarities between the *Drosophila* HOM-C and vertebrate Hox complexes are illustrated (McGinnis & Krumlauf, 1992). HOM-C is shown above, where the broken line between the *Ubx* and *Antp* genes indicates the point where the Antennapedia (ANT-C) and bithorax (BX-C) complexes are separated. The four vertebrate Hox clusters (*HoxA*, *HoxB*, *HoxC* and *HoxD*) are represented below. The genes of each cluster are aligned into their respective paralogous subfamilies (1-13) according to the direction of transcription (5' to 3'). Individual genes are referred to by both the current nomenclature, shown within the box (Scott, 1992), and by the old nomenclature shown below. Note that not all of the clusters have an equal number, or complement, of constituent genes. Gaps within clusters indicate that no known gene exists to fill that position. Paralogous Groups 4 and 9 are the only two subfamilies to have a member from all of the four clusters.

A schematic representation of Hox/HOM gene expression domains within the *Drosophila* embryo and within the central nervous system of the vertebrate (mouse or chicken) embryo are also included (anterior to the right). The spatially colinear relationship between gene order along the chromosome and the relative domain of expression along the AP axis can be seen, such that genes at the 3'-end of the cluster have the most anterior boundaries of expression.

For the *Drosophila* embryo depiction the segments are marked as follows; Int = intercalary, ma = mandibular, Mx = maxillary, Lab = labial, T1-T3 = thoracic, A1-A9 = abdominal. In the schematic of the vertebrate embryo central nervous system, r1-r8 refer to rhombomeric identity within the hindbrain. The dashed oval situated at the axial level of r5/r6 indicates the position of the otic vesicle.
Homeotic Gene Expression Patterns

Drosophila Embryo

Drosophila Genes

Mouse & Chicken Embryos

Mouse & Human Genes

5' - A13 - A11 - A10 - A9 - A7 - A6 - A5 - A4 - A3 - A2 - A1  1.10 - 1.9 - 1.8 - 1.7 - 1.1 - 1.2 - 1.3 - 1.4 - 1.5 - 1.11 - 1.6

5' - B8 - B7 - B6 - B5 - B4 - B3 - B2 - B1  2.5 - 2.4 - 2.3 - 2.2 - 2.1 - 2.7 - 2.8 - 2.9

5' - C13 - C12 - C11 - C10 - C9 - C8 - C6 - C5 - C4  3.7 - 3.6 - 3.2 - 3.1 - 3.3 - 3.4 - 3.5

5' - D13 - D12 - D11 - D10 - D9 - D8 - D4 - D3 - D1  4.8 - 4.7 - 4.6 - 4.5 - 4.4 - 4.3 - 4.2 - 4.1 - 4.9
common identities at the nucleic acid and protein level, both within the homeodomain and in regions external to it, and on the relative positions of the genes within their respective complexes. The vertebrate Hox genes can be grouped into 13 subfamilies or paralogous groups. Members of a given paralogous group (PG) possess shared properties in terms of domains of expression and their location within the Hox clusters. The strongest similarities to the HOM-C genes are seen in PG-1 (lab), PG-2 (pb), PG-4 (Dfd), PG-5 (Scr) and PG-9 through to PG-13 (Abd-B). PG-3 has no Drosophila homologue and the similarities between genes of PG-6 to PG-8 and Antp/Ubx/abd-A make it difficult to draw direct relationships (Krumlauf, 1992, 1994; Kappen & Ruddle, 1993).

Considerable interest has been directed at the mechanisms underlying the evolution of the Hox/HOM complexes. A primordial complex is thought to have arisen by the tandem duplication and divergence of a single progenitor gene early on in eukaryotic evolution (Lewis, 1978; see Duboule, 1994). This would have resulted in a single Hox cluster being present in a common ancient ancestor. Early definition of the nature of this ancestral complex was impeded due to several divergent features of HOM-C. For instance; in Drosophila HOM-C is split into two distinct clusters, ANT-C and BX-C (Lewis, 1978; Kaufman et al., 1980, 1990), the transcriptional orientation of Dfd is opposite to that of all the other homeotic genes, and ANT-C also contains several other non-homeotic genes such as the maternal effect gene bicoid. However, the matter was clarified by the subsequent identification of single, contiguous Hox/HOM cluster in the beetle Tribolium (Beeman et al., 1989) and in various other divergent species (see Krumlauf, 1992, 1994).

Hox clusters from arthropods, vertebrates and nematodes each contain one or more genes related to the Drosophila members lab, pb/Dfd/Scr, Antp/Ubx, Ubx/abd-A and Abd-B (Krumlauf, 1992; Schubert et al., 1993; Wang et al., 1993). This suggests that a common precursor of these taxonomic categories probably contained at least one gene corresponding to each of the four groups (see Kenyon, 1994). Subsequent to the evolutionary divergence of arthropods and vertebrates, independent duplication of certain Hox genes occurred, followed by several replications of the entire Hox cluster somewhere along the vertebrate evolutionary pathway (Kappen et al., 1989; Schubert et al., 1993). As none of the vertebrate Hox clusters contain representative members of all 13 paralogous groups, it is thought that after the duplication of the ancestral Hox cluster, some genes within each cluster became deleted. It is not clear to what extent particular Hox genes have been retained throughout the vertebrate lineage. However, the four-cluster arrangement is a characteristic of all vertebrate genomes so far examined and implies a critical functional requirement for this arrangement. Thus, the expansion of the number of Hox clusters is thought to have been part of a major phase of gene or genome duplication that permitted the evolution of vertebrate development and morphology (Gaunt et al., 1989, 1990; Gaunt, 1991; Holland & Garcia-Fernández, 1996).
1.2.4 Colinearity

One of the most intriguing attributes of the Hox/HOM complexes is a direct linear correlation between the physical location of genes along the chromosome and their domains of expression along the AP axis of the embryo. This characteristic was first perceived in Drosophila for genes of the BX-C and was referred to as colinearity (Lewis, 1978; Harding et al., 1985). The observation that this phenomenon also holds true for the vertebrate Hox clusters provided compelling evidence that Hox and HOM complexes represent structurally and functionally homologous systems (Duboule & Dollé, 1989; Graham et al., 1989).

Experimental studies both in vivo and in vitro have revealed that several different categories of colinearity may be ascribed to the Hox genes. In the embryo, Hox genes exhibit graded patterns of spatially restricted expression that typically exhibit very sharp anterior boundaries, becoming weaker more posteriorly. Spatial colinearity refers to the relationship between the position of a Hox gene within the complex and its anterior limit of expression along the AP axis, such that the most 5' Hox gene has the most posterior boundary and each successive gene has a progressively more anterior one. In vertebrates, for example, which are not overtly segmented, this is most distinct in the neuromeric components of the developing hindbrain (the rhombomeres) which pattern the formation of ganglia and the migration of neural crest into the branchial region of the head (Lumsden, 1990). Within this region the anterior boundaries of expression of the 3' most Hoxb genes (Hoxb-1, -2, -3 and -4) coincide precisely with the segmental boundaries between the rhombomeres. Thus Hoxb-4 has an anterior limit of expression at the rhombomere 6/7 boundary, whereas Hoxb-3 has an anterior limit of expression at the rhombomere 4/5 boundary. One exception to the spatial colinearity rule is seen with Hoxb-1, the most 3' member of the Hoxb cluster, which has an anterior boundary of expression that is more caudal than that of Hoxb-2. This situation is highlighted in Figure 1.3. Unlike Drosophila there is an extensive spatial overlap of Hox expression, so that more posterior regions of the embryo express progressively more Hox genes of a given cluster. Paralogous genes, particularly those at the 3' end of the complex, often display very similar domains of expression in the central nervous system and in the somites. In addition, genes within a given cluster generally show similar tissue-specificities of expression. Spatially colinear Hox gene expression is also seen in other regions of the embryo such as the lateral mesoderm, neural tube, neural crest, limbs, surface ectoderm, branchial arches, gut and gonadal tissue (Gaunt et al., 1988; Dollé et al., 1989, 1991; Duboule & Dollé, 1989; Wilkinson et al., 1989b; Gaunt, 1991; Graham et al., 1991; Hunt et al., 1991b; Kessel & Gruss, 1991). The Hox genes display a colinear relationship with respect to their timing of activation during embryogenesis, termed temporal colinearity (Dollé et al., 1991; Izpisúa-Belmonte et al., 1991a). The colinear properties also extend to the responsiveness of Hox genes to activation by retinoic acid (RA). In both the embryo and in cell culture, Hox genes are differentially sensitive to induction by RA in a concentration and time dependent manner,
Figure 1.3: The Branchial *Hox* Code

The diagram indicates the patterns of gene expression derived from Paralogous Groups 1-4 in the branchial region, after the morphological appearance of rhombomeres and the completion of neural crest migration (after Hunt *et al.*, 1991c). The coloured arrows indicate the migration of mesenchymal and neurogenic crest from specific rhombomeres, resulting in the transference of the combinatorial *Hox* code from the central nervous system into the branchial arches. The branchial arch ectoderm subsequently adopts an identical pattern of paralogous group *Hox* gene expression (indicated by the coloured shading pattern), presumably as a result of interaction with the underlying neural crest. *Hoxb-1* expression is confined to neurogenic (ganglionic) neural crest as indicated by the short red arrow. The large blank arrow emanating from rhombomeres 1 and 2 (r1 and r2) represents the neural crest component of the first branchial arch that does not carry a *Hox* code. Also note the absence of neural crest outflow from r3 and r5 which are crest depleted, by analogy to the chick embryo (Lumsden *et al.*, 1991; Sechrist *et al.*, 1993; see Chapter 1.4.2). The chromosomal relationship between the relevant *Hox* subfamilies expressed in this region is shown below. The large arrow at the bottom of the diagram indicates the direction of transcription and the colinear expression of the genes with respect to boundaries along the AP axis, timing and response to retinoic acid (RA).
such that expression of the 3' most Hox gene is most sensitive to RA and is induced first (Simeone et al., 1990, 1991; Boncinelli et al., 1991; Papalopulu et al., 1991b; Dekker et al., 1992). When the various colinear properties are all related to the organisation of the Hox complex, it can be seen that genes at the 3’ end of the cluster are activated first, have the most anterior boundaries of expression and are most sensitive to RA induction.

1.2.5 Colinearity, clustering and Hox gene regulation

It is thought that the Hox genes themselves represent a molecular code that specifies positional information along the AP axis during development. This is often referred to as the Hox code (Kessel & Gruss, 1991; Hunt & Krumlauf, 1992; McGinnis & Krumlauf, 1992; Krumlauf, 1994). It is well established that *Drosophila* HOM-C genes specify positional information during embryogenesis and their extensive similarity to the vertebrate Hox genes suggests that they have similar functions. Direct evidence for the functional significance of the Hox code, indicating that the precise setting of the boundaries of Hox gene expression is crucial for orderly development, has been derived from; loss- and gain-of-function mutational analyses of the Hox genes which produce characteristic homeotic transformation phenotypes (see Chapter 1.2.6), in vivo modulation of Hox genes by RA (Kessel & Gruss, 1991; Kessel, 1992, 1993; Marshall et al., 1992; Wood et al., 1994), and transplantation experiments (Guthrie et al., 1992).

Despite extensive characterisation of the Hox/HOM complexes the underlying requirement for their striking structural conservation throughout evolution, and the relationship of this to colinearity, remains a mystery. It seems most likely, however, that this has been imposed by functional constraints. It has been postulated that the integrity of the entire Hox cluster is required for the proper regulation and expression of the genes within it. This can be viewed from two extreme standpoints. Firstly, within the Hox complex there exists a single ‘master control region’, located towards the 3’ end of the cluster. This region would be critical for governing the characteristic colinear patterns of Hox gene expression that are observed. A similar type of regulatory mechanism has been well characterised for the expression of the developmentally regulated human β-globin gene cluster (Grosveld et al., 1987; Dillon & Grosveld, 1993). In this case, a combination of stage-specific factors acting at proximal cis-regulatory sequences, and the competition of individual promoters for long-range interaction with the Locus Control Region (LCR), ensures correct sequential developmental gene regulation. The physical interactions between the LCR and individual promoters is thought to occur by a DNA-looping mechanism. The contrasting scenario is that multiple enhancer elements, scattered throughout the length of the Hox cluster, are required for the appropriate regulation of one or several genes. In this type of model the dispersed regulatory elements
would act like 'links of a chain', holding the Hox cluster together. It is possible that the genuine mechanism underlying the coordinated expression of Hox/HOM genes lies somewhere between these two extreme perspectives, involving both long-range and local interactions of cis-regulatory elements acting on single or multiple transcriptional units. Some supporting evidence for this theory may be derived from the results of numerous experiments aimed at recreating the full expression patterns of individual Hox genes in transgenic mice. Employing reporter gene analysis it has been possible to recapitulate the endogenous pattern of gene expression in a few instances, using relatively small genomic regions (approximately 6-18kb), out of their natural genomic context, i.e. for Hoxa-7 (Püschel et al., 1991), Hoxb-4 (Whiting et al., 1991), Hoxb-1 (Marshall et al., 1992) and Hoxa-4 (Behringer et al., 1993). It must be stressed that such reporter gene assays are only valid for the assessment of the crudest aspects of gene expression such as the spatial and, to a lesser extent, temporal distribution of transcripts. As two of the desired properties of a reporter gene assay are heightened sensitivity and stability, no conclusions may be inferred about the precise levels of gene expression nor about the nuances of temporal variation. These results are, however, in accordance with the observation that all of the sequences of the the Drosophila BX-C do not need to be contiguous in order to function properly (Struhl, 1984; Tiong et al., 1987). This tends to argue against a straightforward LCR-based model of Hox gene regulation.

In all of the cases where endogenous Hox gene expression has been faithfully reproduced, reporter constructs include genomic sequences located both 5' and 3' of the transcription start site. It is clear from these limited examples that the mechanisms employed in Hox gene transcriptional regulation are diverse. For example, for both Hoxa-7 and Hoxb-4 multiple regulatory elements located downstream of the transcription start site are required, but have contrasting modes of action. In the case of Hoxa-7, the promoter (5'-flanking DNA) is active in all regions of the embryo and the regulatory elements act to restrict this activity. In Hoxb-4 the promoter is essentially inactive and appropriate expression is only seen in the presence of the regulatory elements. It is also interesting to note that the paralogous genes Hoxa-4 and Hoxb-4 do not share common spatial arrangements of regulatory sequences, since in Hoxa-4 the major regulatory elements appear to be located upstream of the promoter rather than downstream. Hoxa-7 and Hoxb-7 (Vogels et al., 1993) are also dissimilar in two respects. Firstly the regulation of Hoxb-7 is predominantly dependent on activating upstream cis-regulatory sequences, rather than on upstream and downstream lineage-restricting elements. Secondly, 27kb of genomic sequences encompassing the Hoxb-7 gene are unable to recapitulate the full expression pattern of the endogenous gene, whereas for Hoxa-7 approximately 7kb was sufficient. In the majority of instances endogenous Hox gene expression patterns have only been partially reproduced with varying degrees of success, for example; Hoxa-1 and Hoxa-2 (Frasch et al., 1995), Hoxa-5 (Zákány et al., 1988; Tuggle et
Admittedly some of these analyses have only focused on upstream sequences. However, the inability to recreate the endogenous expression pattern of some of the Hox genes suggests that dispersed, and possibly shared, regulatory elements are required. Crucially this line of research has uncovered numerous cis-regulatory elements that are important for particular aspects of Hox gene expression, some of which are capable of acting as spatially-specific enhancers (e.g. Whiting et al., 1991). Thus the identification of trans-acting factors that interact with these sequences will provide important information about how Hox gene expression is activated and maintained in vertebrates.

The presence of interwoven regulatory elements may explain how the Hox/HOM genes have remained clustered but does not explain why this is necessary for ordered, temporally and spatially colinear Hox gene activation. In considering this conundrum we may not be as justified in turning to the Drosophila HOM-C for a comparison with the vertebrate Hox Complex as we have been in other instances. As discussed earlier in this text, many fundamental differences exist between Drosophila and murine embryogenesis at both the molecular and cellular level, particularly with respect to pattern formation. Most evident is the presence of a molecular 'pre-patterning' mechanism involving maternal-effect, gap, pair-rule and segment-polarity genes that predisposes the expression of the HOM-C genes. There is no evidence to suggest that such a system operates in vertebrates, although later phases of Hox/HOM expression appear to utilise analogous mechanisms for the maintenance of the 'on/off' state (Krumlauf, 1994). Some of the genes involved in HOM-C regulation are conserved in vertebrates, and in certain cases are known to be involved in aspects of Hox gene regulation (see Chapter 1.4). However, this does not necessarily mean a directly conserved role in development per se (Lobe & Gruss, 1989). It is possible that the evolution of a pre-patterning system in Drosophila has diverted some degree of evolutionary pressure away from the organisational requirement of HOM-C, particularly in terms of the relative mechanisms for the initial activation of Hox/HOM genes (Kenyon, 1994). This would help to explain some of the more divergent characteristics associated with HOM-C, when compared to Hox clusters on the chordate/vertebrate lineage (Krumlauf, 1994). For example, the split nature of HOM-C and the presence of other non-Hox/HOM like genes that are dispersed amid ANT-C.

Despite the possible differences in the initial mechanisms of activation of Hox/HOM genes, temporal and spatial colinear expression has been described in vertebrates, short germ-band insects such as the locust Schistocerca, and leeches (Krumlauf, 1992, 1994; Akam, 1994; Shankland, 1994). This process is not dependent on spatial cues as it has been observed in cell culture in response to RA treatment, as previously noted. Thus, the key to the conserved
clustered genomic arrangement of \textit{Hox} genes may be to permit ordered temporal expression, resulting in ordered spatial expression. Gaunt \& Singh (1990) have suggested that a bipartite mechanism of transcriptional activation and maintenance could explain the features of temporally and spatially colinear expression of the \textit{Hox} gene complexes. During activation a progressive 3' to 5' opening up of a cluster (i.e. in the same order that the genes are expressed along the AP axis) would result from the activity of stage- and spatially-restricted morphogens on the \textit{cis}-regulatory elements of individual genes. Subsequently the active/inactive state of each gene would be heritably maintained by a silencing mechanism, e.g. by the recruitment of Pc-G-like proteins as previously described. The requirement for gene clustering would therefore be at the level of the maintenance of transcriptional repression. Duboule (1992) proposed a variation on this model which perhaps seems more logical as it places the emphasis for the clustered arrangement at the level of transcriptional activation. This model states: an imperative reason for the conserved clustering of the \textit{Hox} genes might be the need for a carefully controlled sequential activation of the genes from 3' to 5' in the clusters during embryogenesis. \textit{Hox} gene clustering would be required for the fine tuning of the temporal sequence in which the genes are activated (by a still unknown molecular mechanism) during the transmission of positional information in embryos. For this model to hold true, the 'unknown molecular mechanism' would still need to stimulate \textit{Hox} gene activation in a temporal manner, such that the more posterior cells in the embryo activate their \textit{Hox} gene expression later than more anterior ones. In truth of fact, only when we have gained a greater understanding of \textit{Hox} gene expression will we be able to resolve the mechanisms by which colinearity is achieved. Hence the dissection of the regulatory mechanisms governing \textit{Hox} gene regulation is the focus of much research.

1.2.6 Mutational analysis of \textit{Hox} genes in the mouse

In \textit{Drosophila} the genetic function of the \textit{HOM-C} genes was deduced by analysing numerous developmental mutants. Strong correlation between the physical properties of the \textit{Hox/HOM} genes and their segmentally restricted patterns of expression suggested that they belonged to a conserved genetic mechanism (Gaunt, 1991; Kessel \& Gruss, 1991; Hunt \& Krumlauf, 1992). However, direct evidence for genuine functional conservation was not gained until the advent of gene targeting technology, allowing the generation of mice with defined germline mutations (Capecchi, 1989). More recently it has been shown that the chicken \textit{Hoxb-1} protein can functionally substitute for \textit{lab} activity by rescuing a \textit{Drosophila lab} null mutation (Lutz \textit{et al.}, 1996). The presence of homeotic transformations in both dominant gain-of-function and recessive loss-of-function \textit{Hox}-mutant phenotypes suggests that the vertebrate \textit{Hox} genes are also involved in the specification of regional identity along the AP embryonic axis. The general trend of phenotypic alterations correlates with the observed
effects derived from analogous studies in *Drosophila* (see McGinnis & Krumlauf, 1992). Thus, loss-of-function mutations tend to result in anterior transformations, whilst gain-of-function mutations tend to have the opposite effect of posterior transformations. These studies are ongoing and we do not yet have null allelic examples for all 38 murine *Hox* genes. However, targeted mutational analysis of *Hoxa-1, Hoxa-2, Hoxa-3, Hoxa-4, Hoxa-5, Hoxa-6, Hoxa-9, Hoxa-10, Hoxa-11, Hoxb-4, Hoxb-5, Hoxb-6, Hoxc-8, Hoxc-9, Hoxd-3, Hoxd-4, Hoxd-9, Hoxd-10, Hoxd-11, Hoxd-12 and Hoxd-13* generally supports this hypothesis (Chisaka & Capecchi, 1991; Lufkin et al., 1991; Chisaka et al., 1992; Le Mouelllic et al., 1992; Carpenter et al., 1993; Condé & Capecchi, 1993; 1994; Dollé et al., 1993; Gendron-Maguire et al., 1993; Jeannotte et al., 1993; Mark et al., 1993; Ramirez-Solis et al., 1993; Rijli et al., 1993, 1994; Small & Potter, 1993; Davis & Capecchi, 1994; 1996; Horan et al., 1994; 1995a; Kostic & Capecchi, 1994; Rancourt et al., 1995; Suemori et al., 1995 Fromental-Ramain et al., 1996). In each of the above cases the targeted mutation causes regionally restricted defects along the AP or appendicular axes of the mouse. However, the effects of these loss-of-function mutations vary considerably. In some cases, as in the examples of the *Hox* mutations that affect the formation of the vertebral column (e.g. *Hoxa-4, Hoxa-5, Hoxa-6, Hoxa-9, Hoxa-10, Hoxa-11, Hoxb-4, Hoxc-8, Hoxd-3, Hoxd-4, Hoxd-9, Hoxd-11 and Hoxd-13*) or those that effect the formation of the middle ear ossicles (Hoxa-2), the phenotype can be interpreted in terms of homeotic transformations. On the other hand, mutations that affect the formation of the hindbrain (Hoxa-1), the tissues derived from the mesenchymal neural crest (Hoxa-3), or the limbs (Hoxa-9, Hoxa-10, Hoxa-11, Hoxd-9, Hoxd-11 and Hoxd-13), are not readily interpretable as homeotic transformations. Rather, the effects of these mutations are more easily interpreted as the failure to make the affected tissues or the correctly shaped bones, resulting from the lack, or reduced proliferation, of the appropriate precursor cell populations. This phenomenon has been referred to as dysmorphology to distinguish it from true homeosis.

Dominant gain-of-function mutations in the mouse have been achieved with transgenesis by the ectopic expression of coding sequences from a given *Hox* gene, under the control of heterologous transcriptional regulatory sequences. Such mutations have been generated for *Hoxa-1, Hoxa-7, Hoxb-8, Hoxc-6, Hoxc-8 and Hoxd-4* (Balling et al., 1989; Kessel et al., 1990; Jegalian & De Robertis, 1992; Lufkin et al., 1992; Pollock et al., 1992; Charité et al., 1994; Zhang et al., 1994). One approach has been to misexpress one *Hox* gene using the control elements from another. For example, *Hoxd-4* is normally expressed in the paraxial mesoderm with an anterior boundary at the level of prevertebra 1 in the cervical region. When placed under the control of the *Hoxa-1* promoter, which is active in the rostral hindbrain region, *Hoxd-4* is ectopically expressed in more rostral domains, resulting in the homeotic transformation of the occipital bones of the skull towards more posterior identities of the cervical vertebrae (Lufkin et al., 1992). Another approach has been to ubiquitously
express a Hox gene by using a promoter that is not subject to spatial restriction. For example, transgenic mice have been generated that contained the transcribed region of the Hoxa-7 gene under the control of the ubiquitous active β-actin promoter. These animals exhibit abnormalities in the craniofacial region (Balling et al., 1989) and cervical skeleton (Kessel et al., 1990), anterior to the normal domains of Hoxa-7 expression.

It is believed that the Hox genes do not have a critical functional role along their entire axial domain of expression but that a functional hierarchy exists among Hox genes, similar to that seen with the HOM-C genes. This property is termed posterior prevalence in vertebrates and phenotypic suppression in Drosophila. The Hox genes are thought to have a dominant function from the level of their anterior limit of expression, to the anterior limit of more posterior (5' located) Hox genes that are expressed in the same region (González-Reyes et al., 1990; Duboule, 1994). Thus, Hox gene loss-of-function mutations should result in vertebral transformations only at the level of the corresponding anterior domain of expression and not in regions where more posterior genes are expressed. Supporting evidence for this theory has come from the analysis of Hox gene targeted mutations, where the phenotypic effects are concentrated in regions corresponding to a particular gene's anterior boundaries (e.g. Chisaka & Capecchi, 1991; Lufkin et al., 1991; Chisaka et al., 1992; Ramirez-Solis et al., 1993). The results of gain-of-function mutations also support the posterior prevalence model by showing that after the ectopic expression of a Hox gene in a domain more anterior to its normal one, morphological transformations are seen that tend towards the identity of more posterior structures, i.e. those normally expressing the gene (e.g. Kessel et al., 1990; Zhang et al., 1994). In addition, retinoic acid treatment of mouse embryos can result in alterations in the expression domains of Hox genes along the AP axis which also give rise to axial transformations, further supporting the posterior prevalence hypothesis (Kessel & Gruss, 1991). It must be stressed, however, that not all of the phenotypic examples resulting from Hox gene mutations follow this trend. For example, gain-of-function mutations for Hoxc-6 and Hoxc-8 show anterior transformations (Jegalian & De Robertis, 1992; Pollock et al., 1992), whereas loss-of-function mutants for Hoxa-5, Hoxa-6 and Hoxa-11 show posteriorly transformed phenotypes (Jeannotte et al., 1993; Small & Potter, 1993; Kostic & Capecchi, 1994). Similarly, loss-of-function mutations in the Drosophila Dfd and Scr genes result in posterior transformations (Wakimoto & Kaufman, 1981). Further phenotypic analyses will be required to adjudge the nature of such variants.

Hox genes are expressed in a wide range of cell-types and may exhibit dynamic temporal and spatial patterns of expression within a particular tissue (Dollé & Duboule, 1989; Graham et al., 1991; Izpisúa-Belmonte et al., 1991a). Generally, however, the Hox gene mutant phenotypes that have been characterised appear much milder than might have been expected based on expression patterns and by analogy to the HOM-C genes. Members of a given
paralogous group show extensive overlap in their domains of expression and there is a great deal of superposition of Hox gene expression in progressively more posterior regions of the embryo (Gaunt et al., 1989; Hunt et al., 1991a). This has been taken to imply a possible functional redundancy between the Hox genes and may explain the mild phenotypic effects seen with loss-of-function mutations affecting a single gene. The results of Hox gene mutational analyses have provided strong support for the Hox code hypothesis, originally proposed by Kessel & Gruss (1991), and that the specification of segmental identity results from the combinatorial effect of the Hox genes expressed in a given segment. It is not clear, however, whether this combinatorial requirement is effective within all of the same cells of a given segment. Our knowledge of the relative distribution of Hox gene transcripts, and more importantly of Hox proteins, at the cellular level is minimal. Therefore, it is possible that within a given region of the embryo a number of lineage-restricted cellular populations exist, each possessing either a different Hox code or the same code but with different relative proportions of the encoded Hox gene products. Such uncertainties have made the analyses of null mutant phenotypes difficult and have precluded the formulation of a simple, universal Hox code. In light of this, researchers have started to examine the effects of multiple loss-of-function mutations within the same animal and such experiments have begun to yield interesting results. For instance, we now have direct evidence for the existence of partial functional redundancy between genes of the same paralogous group, e.g. Hoxa-3/Hoxd-3, Hoxa-4/Hoxb-4/Hoxd-4 and Hoxa-9/Hoxd-9 (Condie & Capecchi, 1994; Horan et al., 1995b; Fromental-Ramain et al., 1996), but also between non-paralogous genes from the same or separate clusters, e.g. Hoxb-5/Hoxb-6 (Rancourt et al., 1995) and Hoxa-10/Hoxd-11 (Favier et al., 1996). In the case of the PG-3 genes, non-overlapping phenotypic effects were seen with the targeted disruption either Hoxa-3 or Hoxd-3 (Chisaka & Capecchi, 1991; Condie & Capecchi, 1993). Hoxd-3 mice showed a partial anterior homeotic transformation of the first (CV1 or atlas) and second (CV2 or axis) cervical vertebrae, while the compound mutant for Hoxa-3/Hoxd-3 additionally shows a complete loss of the axis (Condie & Capecchi, 1994).

Another potential hurdle for the analysis of loss-of-function mutations is the nature of the mutation itself. Ramirez-Solis et al. (1993) have previously characterised two different null alleles for the Hoxb-4 gene, both of which produced overlapping but different phenotypes. The first null allele (Hoxb-4f) consisted of a disruption of the first exon of the gene by the insertion of two selectable markers. Mice containing this mutation displayed an expected partial anterior homeotic transformation of the identity of CV2 into that of CV1. In addition they also exhibited a defective morphogenesis of the sternum which was unexpected. The second null allele (Hoxb-4t) involved the introduction of a premature translational termination codon into the second exon, disrupting the third helix of the homeodomain and producing a truncated protein. Mice containing this mutation also revealed the partial
transformation of CV2 to CV1 but had no sternum phenotype. Several reasons have been put forward to explain this discrepancy (Ramirez-Solis et al. 1993); (1) The truncated Hoxb-4 protein retains functional properties that allow it to direct correct development of the sternum. It is unlikely that the protein is able to bind to DNA, even at some low level. However, it may be able to interact with other Hox proteins or cofactors to regulate downstream target genes, or cross regulate other Hox genes. (2) The wild-type Hoxb-4 gene may be subject to alternative splicing, producing transcripts that do not require the homeobox and are important for sternum development. (3) The introduction of the selectable marker sequences (approximately 3.1kb in length) into the first exon of Hoxb-4 has had a knock on effect on the regulation of another gene(s) in the cluster, required for sternum development. This type of problem of 'neighbourhood effect' has been seen with Hoxd-10, where its targeted disruption led to the misexpression of the neighbouring Hoxd-9 gene (Rijli et al., 1994), and also with the the closely linked myogenic regulatory genes MRF-4 and Myf-5 (see Olson et al., 1996). It is clear from these examples that for us to fully understand the phenotypic effects observed with loss-of-function mutations in a particular Hox gene, we need to have a very clear understanding of the mechanisms involved in its transcriptional regulation and that of other genes in the same chromosomal region. Once this has been achieved it should be possible to design, with a reasonable degree of confidence, specific mutations within cis-regulatory elements that create a loss-of-function phenotype. This approach would theoretically enable the generation of tissue-specific null alleles, e.g. where the activity of a gene is specifically ablated in the somitic mesoderm.

1.3 The Deformed-related Hox family

The Deformed-related paralogous subfamily of vertebrate Hox genes includes Hoxa-4, Hoxb-4, Hoxc-4 and Hoxd-4, collectively referred to as Paralogous Group 4 (PG-4). This group is one of only two paralogous subfamilies that contain a representative member from each of the four Hox clusters, the other being PG-9 (Featherstone et al., 1988; Graham et al., 1988; Harvey & Melton, 1988; Acampora et al., 1989; Galliot et al., 1989; Sasaki & Kuroiwa, 1990; Sasaki et al., 1990; Geada et al., 1992). Hox genes from PG-1 to -4 have attracted much interest, since the anterior boundaries of their expression domains in the central nervous system coincide with rhombomere boundaries in the embryonic hindbrain. It is thought, therefore, that these genes mediate the branchial Hox code involved in patterning the craniofacial region (Hunt et al., 1991a; Hunt & Krumlauf, 1992).
1.3.1 Characteristics of the Deformed-related *Hox* genes

*Hox* genes are assigned to the *Dfd*-group based on a number of common structural and functional criteria including relative chromosomal position, domains of expression and common sequence identities. The homeodomain of this subfamily is highly conserved between members (approximately 90%) and shares approximately 80% amino acid identity with the archetypal homeodomain of Antp. More importantly, the homeodomains of *Dfd*-group proteins incorporate several idiosyncratic residue substitutions when compared directly to Antp, distinguishing them as 'Dfd-like' (Scott *et al.*, 1989; Bürglin, 1994). Several of the substitutions (i.e. P/S/T/A at positions 1/4/6/7) lie within the flexible N-terminal arm of the homeodomain (see Figure 1.4).

This region makes base specific contacts with residues contained in the minor groove of DNA and is thought to be involved in the determination of target site specificitiy and affinity (Otting *et al.*, 1990; Ekker *et al.*, 1992; Zeng *et al.*, 1993; Pellerin *et al.*, 1994; Phelan *et al.*, 1994; Mann, 1995; Chang *et al.*, 1996). Sequence identity is also seen outside of the homeodomain of *Dfd*-group proteins in two other areas. The first lies at the N-terminus where there is a block of 20-30 conserved amino acids, though no functional role has been assigned to these sequences. The second region, often referred to as the 'extended homeodomain', stretches N-terminally from the homeodomain into the first exon and includes the hexapeptide motif (see Krumlauf, 1992; Mann, 1995). The hexapeptide motif contains a relatively conserved YPWM core sequence and is variably located 5-53 amino acids from the N-terminus of the homeodomain. It is shared by all *Hox* proteins belonging to PG-1 to -8 but is absent in those from PG-9 to -13, i.e. *Hox* genes that are related to *Abd-B* (Acampora *et al.*, 1989). Within the murine PG-4 genes the hexapeptide has the consensus sequence VYPWM[K/R] and is separated by a 12 amino acid linker region of from the N-terminus of the homeodomain (Figure 1.4). The hexapeptide together with residues immediately C-terminal of the homeodomain, which are also conserved between *Dfd*-group proteins, are thought to be important for the interaction of *Hox* proteins with cofactors of the exd/Pbx family (see Chapter 1.4.4) (Mann, 1995; Pöpperl *et al.*, 1995; Shen *et al.*, 1996; Zaho *et al.*, 1996).

In addition to sequence similarities at the protein level, PG-4 *Hox* genes show a large degree of overlap in their domains of expression. In the mouse embryo the major sites of expression are the hindbrain/spinal cord of the central nervous system, somitic mesoderm, and in the derivatives of these tissues (Graham *et al.*, 1988; Galliot *et al.*, 1989; Gaunt *et al.*, 1989; Geada *et al.*, 1992). At approximately 9.0 dpc in the mouse embryo the hindbrain becomes compartmentalised into a number of transient, segmented structures (Lumsden & Keynes, 1989; Lumsden, 1990; Hogan *et al.*, 1994). A number of visible segments form as a series of
neuroepithelial bulges termed rhombomeres (r), which pattern the formation of ganglia and
the migration of neural crest into the branchial region. In the hindbrain Hoxa-4, Hoxb-4 and
Hoxd-4 have anterior boundaries of expression that map to the junction between r6 and r7
(Wilkinson et al., 1989b; Hunt et al., 1991a). Expression domains are evident at
approximately 7.75 to 8.0 dpc, prior to the appearance of morphologically distinct
rhombomeres. At this stage, however, the anterior boundaries of expression are more caudal
and gradually creep forwards to the rostral limit of the presumptive r7 by 8.5 dpc. In the
associated cranial neural crest, these three genes are expressed with an anterior boundary in
derivatives of the fourth branchial arch and in more posterior crest populations of the trunk
(Hunt et al., 1991a). Unlike its three paralogues Hoxc-4 has a more posterior boundary of
expression within the central nervous system and the neural crest. This boundary does not
coincide with a morphologically visible rhombomere interface but is located caudal to the
r6/7 boundary by a distance equivalent to the length of one rhombomere (Geada et al., 1992),
and may therefore map to the junction between r7/8. Despite sharing similar boundaries of
expression along the AP axis, all of the PG-4 Hox gene display notable differences in their
restricted dorsoventral (DV) patterns of expression within the neural tube (Gaunt, 1991;
Graham et al., 1991; Geada et al., 1992). There is a close correspondence of the AP
boundaries of expression in the paraxial mesoderm for all four genes, although the domains
are slightly offset. Thus in the prevertebrae (PV), Hoxd-4 maps to PV1, Hoxb-4 to PV2,
Hoxa-4 to PV3 and Hoxc-4 to PV4 (Gaunt et al., 1989; Whiting et al., 1991; Geada et al.,
1992). These data correlate well with the phenotype of mice carrying loss-of-function
mutations in the Hoxa-4, Hoxb-4 and Hoxd-4 genes, which show partial anterior homeotic
transformations of the cervical vertebrae (CV) corresponding to their anterior boundaries of
expression. Thus, Hoxb-4 +/- and Hoxd-4 +/- mice show a partial transformation of CV2 to CV1
and Hoxa-4 +/- mice display a CV3 to CV2 transformation (Ramirez-Solis et al. 1993; Kostic
& Capecchi, 1994; Horan et al., 1994, 1995a). Compound mutants have also been generated
for Hoxa-4, Hoxb-4 and Hoxd-4. Double mutant combinations show varying degrees of
synergy with respect to the severity of the phenotype obtained. In the triple mutant
combination vertebrae from CV2-C5 are transformed, indicating that the activities of the PG-4
genes are required over a greater AP domain than is revealed by the individual single
mutant phenotypes (Horan et al., 1995b). These results also illustrate that functional
redundancy exists between Hox genes at axial levels caudal to their anterior domains of
expression. In tissues other than the CNS and paraxial mesoderm, the expression patterns of
the four PG-4 Hox genes have not been as well characterised, at least not comparatively. For
Hoxa-4, Hoxb-4 and Hoxd-4 cell-type and stage-dependent differences in the levels and
distribution of expression have been described (Gaunt et al., 1989).
Figure 1.4: Comparison of Dfd-family Hox proteins

(A) Shows a comparison of the 60 amino acid DNA-binding homeodomain sequences from the four murine Paralogous Group 4 Hox proteins (Hoxa-4, Hoxb-4, Hoxc-4 and Hoxd-4), Deformed (Dfd), and the archetypal homeodomain of Antennapedia (Antp). The sequence of the Antp homeodomain is shown in full using the single letter code. Identical residues within the other sequences are denoted by dashes. Protein secondary structure is indicated above the homeodomain sequences, beginning with the N-terminal arm (arm) and followed by three α-helices (Helix 1, 2 and 3/4), which are separated by a loop (l) and a turn (t) structure, respectively. Helical regions of the homeodomains are shown as cylinders, the sequences of which are highlighted in grey. Note the characteristic substitutions of P/S/T/A/Q/V at positions 1/4/6/7/11/13, in all members of the vertebrate Dfd-subfamily.

(B) Shows a comparison of the hexapeptide and sequences separating it from the homeodomain of Dfd-group Hox proteins. The highly conserved hexapeptide sequences are highlighted in grey. Strong sequence similarities are also seen between the linker regions of the proteins, which in the murine proteins are identical in length. Residues that are identical between the Hox sequences are underlined. Numbering is with respect to the first residue if the homeodomain shown in (A). The dots within the murine Hox sequences indicate gaps that have been introduced to obtain maximal alignment with the Dfd linker region which has two additional residues.
A

Antp

Dfd

Hoxa-4

Hoxb-4

Hoxc-4

Hoxd-4

B

Dfd

Hoxa-4

Hoxb-4

Hoxc-4

Hoxd-4
1.4 Regulators of *Hox* gene expression

The extensive similarity between genes of the *Hox/HOM* complexes implies that some aspects of gene regulation may have been conserved throughout evolution. At the same time, however, several significant differences exist. The contribution of cellular position, temporal influences and cell-lineage in establishing appropriate spatiotemporal *Hox/HOM* gene expression may be substantially different from one organism to the next (for reviews see McGinnis & Krumlauf, 1992; Kenyon, 1994; Krumlauf, 1994; Lawrence & Morata, 1994). In *Drosophila* the analyses of numerous developmental mutants has allowed the dissection of the genetic mechanisms involved in the regulation of *HOM-C* gene expression. In the mouse, an important approach has been the identification and characterisation of *cis*-acting regulatory elements that govern restricted aspects of *Hox* gene expression in transgenic mice, using *lacZ* reporter gene analysis (Zákány et al., 1988; Bieberich et al., 1990; Kress et al., 1990; Püschel et al., 1990, 1991; Tuggle et al., 1990; Schughart et al., 1991; Whiting et al., 1991; Jegalian et al., 1992; Renucci et al., 1992; Sham et al., 1992; Behringer et al., 1993; Eid et al., 1993; Gérard et al., 1993; Vogels et al., 1993; Marshall et al., 1994; Studer et al., 1994; Frasch et al., 1995; Knittel et al., 1995; Pöpperl et al., 1995; Shamikant et al., 1995; Becker et al., 1996; Nonchev et al., 1996). The number of *cis*-regulatory sequences that have been defined by this type of approach is steadily increasing. Further characterisation of such elements is expected to lead to the identification of the factors required, either in *cis* or in *trans*, for *Hox* gene regulation. At this stage *Hox* sequences with regulatory potential are rather large and information about precisely defined regulatory units is lacking. With the exception of the interaction between Krox-20 and *Hoxa-2/Hoxb-2* during hindbrain segmentation, as well as the roles of retinoic acid and autoregulatory mechanisms in establishing the rhombomere restricted expression of *Hoxb-1*, no definite links between transcription factors and target sequences have been made (Sham et al., 1993; Marshall et al., 1994; Studer et al., 1994; Pöpperl et al., 1995).

1.4.1 Krox-20

The Krox-20 gene encodes a zinc finger-containing transcription factor that can bind to, and activate transcription from, G-rich target sequences (Chavrier et al., 1988a, 1990; Swimhoff & Milbrandt, 1995). Krox-20 was originally isolated from a fibroblast cell-line as a gene that rapidly responds to serum stimulation (Chavrier et al., 1988b). It belongs to the family of highly related early growth-response (*EGR*) genes and is also known as *EGR-2* (Joseph et al., 1988).

Prior to the morphological appearance of distinct rhombomeres, Krox-20 expression occupies
two stripes that correspond to the presumptive domains of r3 and r5 in the hindbrain of many vertebrate embryos (Wilkinson et al., 1989a; Nieto et al., 1991; Bradley et al., 1992; Oxtoby & Jowett, 1993). Loss-of-function mutations in the mouse Krox-20 gene severely affect the formation of these rhombomeres (Schneider-Maunoury et al., 1993; Swiatek & Gridley, 1993). It is possible to follow the fate of Krox-20 expressing cells in one of the mutants, as the gene was disrupted by an in-frame insertion of a lacZ reporter. In embryos that contain this mutation it appears that r3 and r5 are initially formed but fail to develop correctly and are rapidly eliminated (Schneider-Maunoury et al., 1993). Thus Krox-20 is required for the maintenance of r3 and r5. Recent evidence has indicated that Krox-20 is directly involved in the regulation of Hoxb-2 and Hoxa-2, which are expressed at high levels in r3 and r5 (Sham et al., 1993; Nonchev et al., 1996). Hoxb-2 is expressed in the neural tube with an anterior boundary that coincides with the r2/3 border. During hindbrain segmentation the expression of Hoxb-2 is upregulated in r3, r4 and r5. The enhancer that mediates this aspect of Hoxb-2 expression has been mapped by transgenic analyses and can be separated into two elements that control either r4 or r3/5 activity (Sham et al., 1992, 1993). The r3/5 enhancer contains three in vitro binding sites for Krox-20 which are essential for r3/5 upregulation. Furthermore, ectopic expression of Krox-20 activates a reporter gene carrying the Hoxb-2 r3/5 enhancer in transgenic mice. In Krox-20 heterozygous mutant embryos Hoxb-2 expression is initially seen in the hindbrain but is specifically lost in r3 and r5, prior to their degeneration. These findings apparently confirm that Hoxb-2 is a direct target for regulation by Krox-20 within r3 and r5 of the mouse hindbrain. Krox-20 activity only seems to be required for modulation of Hoxb-2 expression, however, and not for its initial activation. A similar situation is seen with the Hoxa-2 gene which has an anterior boundary of expression in the hindbrain one rhombomere rostral to that of Hoxb-2, at the r1/2 border (Hunt et al., 1991a; Krumlauf, 1993). Upregulation of Hoxa-2 in r3/5 is also dependent on a similar upstream enhancer, containing Krox-20 binding sites that are essential for its activity (Nonchev et al., 1996).

Krox-20 may have a role in the regulation of other Hox genes, both in the hindbrain and elsewhere in the embryo. The three constituent genes of PG-3; Hoxa-3, Hoxb-3 and Hoxd-3, are all expressed in the hindbrain with common anterior boundaries that coincide with the r4/5 border. Hoxa-3 and Hoxd-3 show high levels of expression in r5 and could be additional targets for Krox-20. Interestingly, perhaps, the first direct association between Krox-20 binding to a Hox gene regulatory region was reported as an in vitro observation with the Hoxa-4 5'-flanking region, to which Krox-20 is able to bind to and activate transcription from (Chavrier et al., 1990). It was recognised, however, that the expression domains of these two genes do not overlap at early stages of mouse embryogenesis. In this study we too have identified a similar G-rich binding site, which may also bind Krox-20 or a related factor, and is important for the activity of a major Hoxb-4 enhancer in transgenic mice. More recent
analysis of Krox-20+ mice has shown that the gene is also involved in the processes of bone formation, and myelination within the peripheral nervous system (PNS) (Topilko et al., 1994; Levi et al., 1996). In the PNS, Krox-20 is expressed in the boundary cap cells that surround nerve exit points from the central nervous system. Between 10.5-14.5 dpc expression is confined to the motor and sensory roots of cranial and spinal nerves (Topilko et al., 1994). This pattern of expression does overlap with that of Hoxb-4 (Whiting et al., 1991). It will be interesting to see if Krox-20 also plays a role in the regulation of Hox gene expression during these latter stages of embryogenesis.

1.4.2 Kreisler

A mutant mouse strain named kreisler (kr) has been described which exhibits defects in the development of the inner ear (Frohman et al., 1993; McKay et al., 1994). The kreisler phenotype is caused by a recessive mutation that was induced by X-ray exposure. Recently the gene involved has been cloned and was found to encode a novel transcription factor of the bZIP class (Cordes & Barsh, 1994).

Homozygous kr/kr embryos have defects in hindbrain segmentation related to the loss, or transformation, of r5 (Cordes & Barsh, 1994; McKay et al., 1994). Morphological rhombomere boundaries posterior to that of r3/4 are absent, resulting in a smooth thickening of the posterior hindbrain neuroepithelium. In the chick embryo, fate-mapping studies have demonstrated that whilst rhombomeres 1/2, 4 and 6 are foci of neural crest production, the intervening rhombomeres, r3 and r5, are extensively depleted of crest (Lumsden et al., 1991; Sechrist et al., 1993). The loss of neural crest precursors from r3 and r5 involves the induction of Bmp-4 and msx-2 expression and their subsequent apoptotic elimination (Graham et al., 1994). It is thought that this process sculpts the neural crest outflow from r1/2, r4 and r6 into three distinct streams, each of which populate a different branchial arch (see Figure 1.3). It has been shown that the specific depletion of crest from r3 and r5 does not occur in an autonomous manner, but is under the interactive control of the intervening even-numbered rhombomeres (Graham et al., 1993, 1994). Neural crest migration in kr/kr mutant embryos follows two pathways; the first migrates laterally from r2 as is the case in wild-type embryos, the second migrates in a broadened domain from the region of r4 to r7 to populate both the second and third branchial arches. It has been suggested that with the loss of r5 and its associated crest-depleted zone, neural crest from r4 and r6 could mix freely. Developmental abnormalities have been observed in kr/kr animals, such that structures normally associated with the second branchial arch are associated with the third (Frohman et al., 1993). The expression patterns of Hoxb-1, Hoxb-3 and Hoxb-4 are altered in kr/kr embryos in a manner consistent with the loss of r5. In the case of Hoxb-4, its anterior limit of
expression is shifted rostrally by a distance equivalent to one rhombomere length (Frohman et al., 1993). In addition, the normal r5 domain of Krox-20 expression was never observed in mutant embryos. This would suggest that kreasler activity acts upstream of Krox-20 and could explain the absence of morphological rhombomere boundaries. It is thought that the kr/kr phenotype may stem from the neuroepithelial cells in r5 and r6 adopting a r4 and r7 identity, respectively. Whether or not kreasler acts as a direct regulator of Hox gene expression remains to be seen.

1.4.3 Retinoic acid

Exposure to increased levels of retinoic acid (RA) has been shown to alter the expression of Hox genes in cell-lines (Simeone et al., 1990, 1991; Papalopulu et al., 1991b) and embryonic tissues (Izpisúa-Belmonte et al., 1991b; Kessel & Gruss, 1991; Morriss-Kay et al., 1991; Papalopulu et al., 1991a; Conlon & Rossant, 1992; Marshall et al., 1992; Wood et al., 1994). Hox genes exhibit a colinear response to RA, such that genes located most 3' in the clusters respond earlier and to lower concentrations of RA than successively more 5' genes (Simeone et al., 1990), which may be unresponsive to RA or even repressed by it (Papalopulu et al., 1991b; Simeone et al., 1991). Within cultured cell-lines the transition between responsive and unresponsive genes roughly correlates with those of PG-9 (see Figure 1.2) (Boncinelli et al., 1991). In embryological terms, the axial level at which this transition between the two sets of Hox genes occurs coincides with the thoraco-lumbar boundary. This boundary also marks the changeover in the recruitment of paraxial mesoderm from the primitive streak to the tail bud (Tam & Tan, 1992).

In the mouse embryo RA exposure has been shown to rapidly induce anterior shifts in the expression domains of the 3' Hoxb genes (Hoxb-1, Hoxb-2, Hoxb-4 and Hoxb-5) in the neural tube (Morriss-Kay et al., 1991; Conlon & Rossant, 1992; Marshall et al., 1992). No change was observed with the most 5' gene of the cluster (Hoxb-9), which is also not RA-inducible in cultured cells (Conlon & Rossant, 1992). RA-induced anterior shifts of Hox gene expression also correlate with posterior transformations of the vertebrae (Kessel & Gruss, 1991). However, if similar RA treatment is administered at a later stage of development when more caudal somites are forming, anterior transformations are observed (Kessel, 1992). The results of these experiments suggest that RA influences the patterning of the central nervous system and paraxial mesoderm. It is not known whether the widespread effect on Hox gene expression domains results from a direct effect of RA (acting through its nuclear receptors) regulating Hox gene expression, or whether RA acts through an indirect mechanism.

The main effects of retinoic acid are mediated through nuclear receptors belonging to the
superfamily of ligand-activated nuclear transcription factors. Two classes of retinoid receptors have been identified (Glass et al., 1991; Linney, 1992; Stunnenberg, 1993; Chambon, 1994), both of which exist in three major forms, encoded by separate genes each of which generates multiple isoforms; the retinoic acid receptors (RAR) \( \alpha \), \( \beta \) and \( \gamma \) bind both all-trans-RA and its isomer 9-cis-RA, whereas the retinoid X receptors (RXR) \( \alpha \), \( \beta \) and \( \gamma \) bind 9-cis-RA but do not bind all-trans-RA with high affinity (Heyman et al., 1992; Leid, 1994; Mangelsdorf et al., 1994). RXRs can heterodimerise with a number of nuclear receptors including the RARs. RARs need RXRs as heterodimerisation partners for effective transcriptional activation of target genes (Kliewer et al., 1992b; Leid et al., 1992b), whereas RXRs can form active homodimers (Zhang et al., 1992). RAR and RXR genes are widely and differentially expressed throughout the embryo (Dollé et al., 1990; Ruberte et al., 1990, 1991a, 1991b; Mangelsdorf et al., 1992; Dollé et al., 1994) and their protein dimerisation products bind to specific DNA sequence motifs termed retinoic acid response elements (RARE), thereby affecting the transcription of target genes. The canonical RARE consists of two directly repeated (DR) half-sites, separated by a variable number of spacer nucleotides of the consensus PuG[G/T]TCA(X)\( _n \)PuG[G/T]TCA, where Pu denotes a purine nucleotide (A or G) and \( (X)_n \) is the number bases that separate the repeats (Leid et al., 1992a; Stunnenberg, 1993). Transactivation and DNA binding studies using artificial response elements have suggested that the exact DR spacing may determine the specificities of dimeric receptor binding. Thus, RAR/RXR heterodimers bind to elements with spacers of 1, 2 or 5bp (DR\(_1\), DR\(_2\) and DR\(_3\)), whilst interactions with thyroid hormone (TR/RXR) and vitamin D (VDR/RXR) receptors favour DR\(_4\) and DR\(_3\) sites, respectively. RXR/RXR homodimers bind specifically to DR\(_1\) elements. More recently it has been demonstrated that widely spaced half-sites, up to 150bp apart, are also able to act as response elements for a range of different nuclear receptors. For example, RAR/RXR heterodimers show maximal activation through a DR\(_{15}\) element, with a DR\(_{50}\) element being equivalent to the classical DR\(_5\) (Umesono et al., 1991; Kliewer et al., 1992b; Kato et al., 1995).

Despite the wide ranging effects of ectopic exposure to RA on the expression patterns of *Hox* genes, little is known about how this effect is generated. Studies on the *Hoxa-1* gene were the first to reveal a functional RA-responsive enhancer among the vertebrate *Hox* genes (Langston & Gudas, 1992). This enhancer, located downstream of the *Hoxa-1* gene, was found to contain a functional DR\(_5\)-RARE motif essential for the RA-response of *Hoxa-1* in cell culture. More recently the importance of this element for the in vivo expression of *Hoxa-1* has been demonstrated in transgenic mice (Frasch et al., 1995). Rhombomere restricted expression of *Hoxb-1* is partly governed by the activity of two RARE-dependent enhancer elements that have contrasting modes of action. A 3' neural enhancer is required for the early activation of *Hoxb-1* in the presumptive hindbrain. Subsequently the expression domain of
*Hoxb-1* becomes restricted to r4, requiring the activity of a negatively acting 5' element which specifically represses the expression of *Hoxb-1* in r3 and r5. The action of both of these regulatory elements depends on the presence of functional DR₂-RARE motifs (Marshall et al., 1992; Studer et al., 1994). The repressive role of the *Hoxb-1* 5' DR₂-RARE is unusual as RAR/RXR activity is generally implicated in positive regulation. It has been suggested that likely candidate cofactors for the modulation of RAR and RXR activities are members of a different class of nuclear transcription factors; namely the orphan nuclear receptors, Chicken Ovalbumin Upstream Promoter Transcription Factors (COUP-TF) (Wang et al., 1989). Highly conserved members of this family have been isolated from a range of vertebrate species, that in mouse includes the proteins encoded by the *COUP-TFI* and *COUP-TFII* genes (Qiu et al., 1994). A *Drosophila* homologue of COUP-TF is encoded by *seven-up* (*svp*), required for neurogenesis and photoreceptor development (Mlodzik et al., 1990b). COUP-TFs have been shown in cell culture to negatively regulate RAR/RXR and RXR/RXR activities on a large variety of RAREs, either by direct binding-site competition between COUP-TFs and retinoid receptors and/or the sequestration of RXR into inactive COUP-TF/RXR heterodimers (Cooney et al., 1992, 1993; Kliewer et al., 1992a; Tran et al., 1992; Widom et al., 1992; Lu et al., 1994).

Another case of a functional DR₅-RARE regulating *Hox* gene expression in cell culture has been described for the *Hoxd-4* gene (Pöpperl & Featherstone, 1993). This element is located approximately 2.4kb upstream of the *Hoxd-4* coding region and is required for the RA-induced expression of a *Hoxd-4* reporter construct in embryonal carcinoma cells. Furthermore, it was shown that the response of *Hoxd-4* to RA could be inhibited by the expression of a truncated form of RARα that acts in a dominant-negative fashion. Although the significance of this finding in an embryonic context is unknown, it may suggest that RA-responsive enhancers are also involved in regulating the expression patterns of PG-4 *Hox* genes. Interestingly perhaps, the anterior boundaries of *Hoxa-4*, *Hoxb-4* and *Hoxd-4* in the mouse hindbrain coincide with that of RARβ (Ruberte, 1991a; Whiting et al., 1991). It has also been shown that mouse mutants for the RARγ gene have homeotic phenotypes, some of which resemble those observed in *Hoxb-4* mutants (Lohnes et al., 1993; Ramirez-Solis et al., 1993).

### 1.4.4 *Hox* genes and cofactors

As previously discussed in Chapter 1.2.2, autoregulatory and crossregulatory interactions among the *HOM-C* genes, involving both direct and indirect mechanisms, play an important role in the establishment of their final expression patterns. The autoregulatory mechanisms
of the Dfd gene are among the most extensively characterised (Kuziora & McGinnis, 1988; Regulski et al., 1991; Zeng et al., 1994; Lou et al., 1995). There is also evidence to suggest that similar modes of regulation may also feature in the expression of the murine Dfd-group Hox genes. For the Hoxd-4 gene, an upstream enhancer element with autoregulatory potential has been identified in cultured cells (Pöppelr & Featherstone, 1992). In vitro, the Hoxa-4 protein is capable of interacting with several homeodomain binding-sites located within the Hoxa-4 intron and upstream region (Wu & Wolgemuth, 1993). Perhaps most provocatively though, Hoxb-4 mutant mice exhibit a caudal shift in the anterior boundary of Hoxb-4 expression in the central nervous system, suggesting a requirement for Hoxb-4 protein function (Ramirez-Solis et al., 1993).

Recent research has shown that r4 restricted Hoxb-1 expression in the hindbrain is dependent on an autoregulatory loop, requiring the activity of cofactors of the exd/Pbx family (Pöppelr et al., 1995). As mentioned earlier, Hoxb-1 is initially expressed throughout the presumptive hindbrain region and later becomes restricted to r4, requiring the activity of both positive and negative elements (Marshall et al., 1992; Studer et al., 1994). In transgenic mice the ectopic expression of Hoxa-1 or Hoxb-1 under the control of the β-actin promoter can induce Hoxb-1 expression in r2 (Zhang et al., 1994). This effect has also been observed as a response of Hoxb-1 to RA which requires the activity of a 3' RARE (Marshall et al., 1992). An upstream enhancer element was found to be responsible for mediating the auto-induction of Hoxb-1 expression by lab-family members. The activity of this enhancer is dependent on three conserved repeat-motifs to which the Hoxb-1 protein can bind in vitro, but only in the presence of a cofactor. Using parallel reporter gene studies in Drosophila, where the Hoxb-1 r4 enhancer is functional, it was demonstrated that the activities of both lab and the gene extradenticle (exd) are required. Exd encodes a homeodomain-containing protein that can function as a cofactor for HOM-C gene activity and is a homologue of the human proto-oncogene Pbx1 (Rauskolb et al., 1993; Rauskolb & Wieschaus, 1994). It was further shown in vitro that exd/Pbx1 could cooperatively interact with Hoxb-1 and bind to the r4 enhancer repeat-motifs (Pöppelr et al., 1995). The results of these investigations suggest that Hoxb-1 expression in r4 is controlled by a direct, cofactor-dependent autoregulatory mechanism requiring the gene products of both lab and exd/Pbx family members.

These finding have important implications, not only for the elucidation of similar mechanisms that may regulate other Hox genes, but also for the involvement of Hox proteins in the regulation of downstream targets. It is thought that Hox/HOM proteins control regionalised cell fate by regulating the transcription of different, though partly overlapping, sets of downstream target genes (Andrew & Scott, 1992; Botas, 1993). However, in vitro most Hox/HOM proteins bind to the same or very similar binding sites containing a TAAT core with similar affinities (Beachy et al., 1988; Affolter et al., 1990; Ekker et al., 1991;
To account for this it has been suggested that there must be cofactors that modify the DNA-binding and/or regulatory properties of Hox/HOM proteins (Gehring et al., 1994; Krumlauf, 1994; Lawrence & Morata, 1994). The discovery of the role of proteins belonging to the exd/Pbx family provides direct evidence that this is the case, refining the repertoire of potential Hox/HOM target sequences (Mann, 1995). In vitro binding-site selection analysis, aimed at determining the sequence of high-affinity target sites recognised by a variety of Hox-Pbx heterodimeric complexes, has defined a preferred consensus (Chang et al., 1996). The sequence of this consensus (5' - ATGATTNNATNN-3') consists of two adjacent half sites in which the Pbx component of the heterodimer contacts the 5' half (ATGAT), and the Hox component contacts the more variable 3' half (TNATNN).

Three mammalian Pbx genes have been so far identified; Pbx1, Pbx2 and Pbx3 (Kamps et al., 1990; Monica et al., 1991). In vitro analysis has demonstrated that Pbx proteins are able to heterodimerically bind to DNA with a variety of Hox proteins, including Hoxb-4 (van Dijk et al., 1995; Shen et al., 1996). This cooperativity is thought result from combined interactions of the homeodomain and YPWM motifs (Mann, 1995; Neuteboom et al., 1995; Shen et al., 1996). The linker region between these two motifs may also retain some function, particularly with respect to the stability of complex formation. Thus, the protein products of the PG-4 genes, which have almost identical homeodomain, linker and hexapeptide sequences, form binding complexes with Pbx and DNA that have very similar half-lives (Shen et al., 1996). Hox/HOM proteins that lack a recognisable hexapeptide motif also appear to be able to interact with exd/Pbx and bind to DNA, suggesting that the C-terminal region of the homeodomain may also play a role (Chan et al., 1994; van Dijk & Murre, 1994; Chang et al., 1995b). If cofactors such as exd/Pbx truly have such a vital function in determining Hox/HOM target specificity, then it follows that their role in pattern formation is as almost as significant as that of Hox/HOM genes themselves.

1.4.5 Caudal/Cdx

Members of the caudal-type family may play a direct role in regulating early events of Hox gene expression. The Drosophila caudal (cad) gene encodes a divergent homeodomain-containing protein with both maternal and zygotic components. The maternal products of cad form a gradient that peaks at the posterior pole of the embryo (Mlodzik et al., 1985; Mlodzik & Gehring, 1987). This is in contrast to the products of the gap gene hb that form an opposing gradient (Tautz, 1988). Two promoters, which drive cad expression from a single protein coding region, are specifically employed for either maternal or zygotic transcription (Mlodzik & Gehring, 1987). In embryos that lack maternal cad there is some
disturbance of the global AP pattern (Macdonald & Struhl, 1986). The maternal *cad* product directly activates zygotic *fitz* transcription in the posterior half of the embryo (Dearolf et al., 1989). Zygotic *cad* expression is confined to a single posterior band in parasegment 15 where it appears to act as the main patterning determinant. Later in development the gene is persistently expressed in the hindgut, posterior midgut and Malphigian tubules through to adult life (Mlodzik et al., 1985; Mlodzik & Gehring, 1987). Animals mutant for zygotic *cad* are missing structures from the eighth abdominal segment, as well as more posterior segments, but do not exhibit homeotic transformations (Macdonald & Struhl, 1986). Ectopic expression of *cad* in the anterior end of the cellular blastoderm has been found to disrupt head development and segmentation owing to the altered expression of *fitz* and *engrailed* as well as the repression of head-determining *HOM-C* genes, such as *Dfd* (Mlodzik et al., 1990a).

*Caudal*-like genes have been identified in a number of diverse species such as hamster, chicken, *Xenopus laevis*, zebra fish, and *Caenorhabditis elegans* (see Subramanian et al., 1995). Three such genes have been isolated in the mouse; *Cdx1*, *Cdx2* and *Cdx4* (Duprey et al., 1988; James & Kazenwadel, 1991; Gamer & Wright, 1993; James et al., 1994; Suh et al., 1994). All of these genes are expressed during early gastrulation, and at later stages *Cdx1* and *Cdx2* are expressed in the intestinal epithelium (Gamer & Wright, 1993; Meyer & Gruss, 1993; James et al., 1994; Subramanian et al., 1995). The embryonic activity of *Cdx1* has been determined in detail at both the mRNA and protein levels, suggesting a role in the the specification of positional information along the AP axis (Meyer & Gruss, 1993). *Cdx1* mRNA and protein are first observed at 7.5 dpc in the the ectoderm and mesoderm of the primitive streak in a graded fashion, with the highest levels at the posterior end of the embryo. Expression within the neuroectoderm initially extends along the axis of the embryo, posterior to the presumptive mid-hindbrain region, regressing to the level of the spinal cord by 9.5 dpc. A similar progression of mesodermal expression is also seen, where initially all of the somites express *Cdx1* followed by a posterior shift to the level of the spinal cord. *Cdx1* expression is strongly down-regulated after 10.5 dpc but persists along the dorsal somitic edge, in the developing kidney, and the limb until 12 dpc.

Recently the murine *Cdx1* gene has been inactivated by targeted disruption (Subramanian et al., 1995). *Cdx1*−/− mutant mice are viable but show widespread anterior homeotic transformations of the vertebrae. These transformations were shown to be concomitant with posterior shifts in the domains of *Hox* gene expression within the paraxial mesoderm, indicating that *Cdx1* acts as a regulator of *Hox* genes in this tissue. Although the domain of *Hoxb-4* expression was not analysed, the anterior boundaries of *Hoxd-3*, *Hoxc-5*, *Hoxc-6*, *Hoxa-7* and *Hoxc-8* were all found to be shifted posteriorly by a distance of one or two segmental units. Furthermore, the phenotypic partial transformation of the second cervical
vertebra into the first cervical vertebra closely resembles that seen in *Hoxb-4*⁻⁻ mutant animals (Ramirez-Solis *et al.*, 1993). Taken together these observations suggest that *Hoxb-4* expression may also be affected. *Cdx1* only seems to be required for establishing the appropriate anterior domains of *Hox* gene expression in the somites, since no neural defects were observed and the hindbrain boundary of *Hoxd-3* was unaltered by the mutation.

It is likely that *Cdxl* influences transcription directly by binding to *Hox* gene *cis*-regulatory elements. Subramanian *et al.* (1995) have identified two putative Cdx1 binding motifs within an enhancer element of the *Hoxa-7* gene, through which Cdx1 can *trans*-activate a reporter gene in cultured cells. The binding site requirement for chicken CdxA has been closely examined, allowing the consensus binding site [A/C]TTTTAT[A/G] to be determined (Margalit *et al.*, 1993). *Cad* family members have been implicated in the regulation of *Hoxc-8* (Shashikant *et al.*, 1995). In addition, *Cdx4* has been cloned by virtue of its involvement in the regulation of the sucrase-isomaltase (SI) gene, where it acts by binding to a conserved element within the SI promoter (Suh *et al.*, 1994). It is likely that the other murine *cad* family members may have a role that is similar to, and possibly overlapping with, that of *Cdx1*, since *Cdx2* and *Cdx 4* are both active early on in gastrulation and their domains of expression partially overlap with that of *Cdx1* in the posterior part of the embryo (Gamer & Wright, 1993; Subramanian *et al.*, 1995).

### 1.4.6 Polycomb and trithorax related factors

As previously described in Chapter 1.2.2, genes of the *Drosophila Polycomb* (*Pc-G*) and *trithorax* (*trx-G*) groups play a crucial role in the maintenance of the repressed or active state of the *HOM-C* genes, after their early domains of expression have been established (Paro, 1993). They are thought to function by the modification of chromatin structure. Recently mammalian counterparts related to the *Pc-G* genes have been discovered, which include *mel-18, bmi-1* and *M33* (Tagawa *et al.*, 1990; Brunk *et al.*, 1991; Pearce *et al.*, 1992). The *mel-18* and *bmi-1* gene products share amino acid sequence homology with proteins encoded by the *Drosophila Pc-G* genes *Posterior sex combs* (*Psc*) and *Suppressor 2 of Zeste* (*Su(z)2*) (Brunk *et al.*, 1991; Haupt *et al.*, 1991; van Lohuizen *et al.*, 1991; Ishida *et al.*, 1993). The presence of *Pc-G* and *trx-G* homologues in vertebrates suggest that common mechanisms also exist for the regulation of *Hox* gene expression (see Krumlauf, 1994).

The oncogene *bmi-1* was originally identified due to its involvement in the formation of B- and T-cell lymphomas (Brunk *et al.*, 1991; Haupt *et al.*, 1991; van Lohuizen *et al.*, 1991). The products of the *Pc-G* related genes *mel-18, bmi-1, Psc* and *Su(z)2* share amino acid sequence and secondary structure, including a RING-finger motif (Freemont *et al.*, 1991). As
previously mentioned, products of some of the Pc-G genes (e.g. \textit{Pc} and \textit{ph}) do not seem to bind to DNA in a sequence-specific manner (see Chapter 1.2.2). Recently, however, it has been demonstrated that the mel-18 protein is a sequence-specific DNA binding protein and acts to negatively regulate transcription \textit{via} its cognate binding sequence GACTNGACT (Kanno \textit{et al}., 1995). Mice lacking the \textit{bmi-1} gene show, in addition to neurological and hematopoietic defects, posterior transformations of the axial skeleton (van der Lugt \textit{et al}., 1994). In contrast, the over-expression of \textit{bmi-1} in transgenic mice has the opposite effect, causing anterior transformations of the vertebrae in a dose-dependent manner (Alkema \textit{et al}., 1995). At 12.5 dpc \textit{bmi-1} is ubiquitously expressed throughout the embryo with elevated levels in the nervous system, kidney and lung. The effect of over-expressing \textit{bmi-1} was not to drastically alter the distribution of \textit{bmi-1} in the embryo, but to greatly enhanced the levels of expression. Within the vertebral column, abnormalities were most evident in the cervical region and mimicked the effect seen in \textit{Hoxb-4}\textsuperscript{-/-} mice, i.e a partial transformation of CV2 into CV1 (Ramirez-Solis \textit{et al}., 1993; Alkema \textit{et al}., 1995). Similar anterior transformations were also observed in the spinal ganglia. Anterior homeotic transformations were apparent along the entire AP axis of the axial skeleton as was the case in \textit{bmi-1}\textsuperscript{-/-} mutant mice, although in the opposing direction (van der Lugt \textit{et al}., 1994). The presence of widespread vertebral transformations in both dominant and null \textit{bmi-1} mutant animals suggests that the gene is involved in the repression of multiple \textit{Hox} genes. Van der Lugt \textit{et al}., (1994) observed that the anterior boundary of \textit{Hoxc-5} expression in the prevertebrae (PV) was shifted posteriorly from the level of PV5/6 to PV6/7 in dominant \textit{bmi-1} mutant embryos. In \textit{Drosophila}, loss-of-function mutations in Pc-G genes have no effect on the initial expression domains of the \textit{HOM-C} genes. However, at later stages of embryogenesis they become ectopically expressed (Simon \textit{et al}., 1992; Martin & Adler, 1993). Since Pc-G genes are also involved in the regulation of gap gene expression earlier in \textit{Drosophila} development, detailed analysis of \textit{Hox} gene expression patterns will be required to determine at what stage \textit{bmi-1} is required.

Genes related to \textit{trithorax} (\textit{trx}) and \textit{brahma} (\textit{brm}) which act as positive regulators of \textit{HOM-C} expression, presumably by counterbalancing the effects of Pc-G activity, have also been characterised in vertebrates. The human mixed-lineage leukemia gene (\textit{MLL, HRX, ALL-1}) was cloned by virtue of its involvement in chromosomal translocations associated with acute childhood leukaemia (Djabali \textit{et al}., 1992; Gu \textit{et al}., 1992; Tkachuk \textit{et al}., 1992; Rowley, 1993). The murine homologue of this gene (\textit{Mll, All-1}) shares over 90% identity with the human protein and contains three domains that exhibit sequence conservation with \textit{Drosophila} \textit{trx} (Ma \textit{et al}., 1993). These include: (1) Two cysteine-rich regions within the middle of the protein that are believed to be involved in the formation of DNA-binding zinc fingers. In \textit{trx} this domain is essential for \textit{trx}-dependent activation (Chang \textit{et al}., 1995a); (2) The C-terminal region containing a highly conserved SET domain also found in Pc-G
proteins (Simon, 1995); (3) An N-terminal 'AT-hook' domain that is also found in high mobility group proteins, where it is involved in binding to the minor groove of AT-rich DNA (Reeves & Nissen, 1990). Several mammalian genes have been cloned which show similarity to the *Drosophila* trx-G member *Brahma* (*brm*) (Kennison & Tamkun, 1988; Okabe *et al.*, 1992; Soininen *et al.*, 1992; Tamkun *et al.*, 1992; Randazzo *et al.*, 1994; see Chapter 1.2.2). The most related of these and putative homologues of *brm* are the murine *brgl* and human *BRG1* genes (Khavari *et al.*, 1993; Randazzo *et al.*, 1994). The proteins encoded by *brm*/brgl show strong homology to the yeast transcriptional activator SNF2/SWI2 which possesses DNA-stimulated ATPase activity. In yeast SNF2/SWI2 forms part of a large multisubunit complex that can prevent histone-mediated repression, and act as a global activator of promoters that require dynamic remodelling of chromatin structure for transcription (Travers, 1992; Yoshinaga *et al.*, 1992; Becker, 1994; Cairns *et al.*, 1994). The greatest degree of sequence conservation between *brgl*, *brm* and SNF2/SWI2 is seen in the putative helicase domain and in the bromodomain (Laurent *et al.*, 1993; Randazzo *et al.*, 1994). Hence it thought that *brm*/brgl may play a similar role as activators of *Hox/HOM-C* expression. In the mouse, *brgl* is expressed ubiquitously from at least 7.5 dpc onwards, showing the highest levels within the developing central and peripheral nervous systems and in the vertebral column, consistent with a role in *Hox* gene regulation (Randazzo *et al.*, 1994).

Recently the mouse *Mll* gene has been disrupted by gene targeting (Yu *et al.*, 1995). *Mll* is widely expressed from at least 7.0 dpc in the embryo and persists into adulthood. *Mll*+/− heterozygotes show retarded growth, hematopoietic abnormalities and exhibit bidirectional homeotic transformations of the axial skeleton as well as sternal malformations. In agreement with this phenotype, the anterior boundaries of mesodermal and neural expression of the representative *Hox* genes, *Hoxa-7* and *Hoxc-9*, are shifted posteriorly by a distance of one to three segmental units. These effects on *Hox* gene expression are observed as early as 9.5 dpc. The homozygous state (*Mll*−−) is embryonic lethal and embryos die midway through gestation. At 10.5 dpc expression of *Hoxa-7* and *Hoxc-9* was undetectable, despite the presence of normal sites of *Hox* expression such as somites, neural tube, spinal ganglia and limb buds. The results of these experiments firmly establish a positive regulatory role for *Mll* in *Hox* gene regulation. However, it is not clear whether *Mll* is required for the initial activation or for the maintenance of *Hox* gene expression, or both. A similar effect of phenotypic dosage-sensitivity is seen with both *Mll* deficiency and *bmi-1* over-expression (van der Lugt *et al.*, 1994). This provides further evidence that Pc-G and trx-G proteins operate by a mechanism of competitive interaction, as has been suggested for *Drosophila* (Chang *et al.*, 1995a; see Chapter 1.2.2). In this respect it would be interesting to see if *bmi-1*+/−/Mll+/− double mutants exhibit a less severe phenotype than with either mutant alone, as has been observed for *Drosophila trx/Pc−* double mutants (Capdevila & Garcia-Bellido, 1981; Sato & Denell, 1987). The results of mutational analyses of *Cdx1*, *bmi-1* and *Mll*, which
highlight the correlative effects between the alteration of anterior boundaries of Hox gene expression and the production of vertebral transformations, provide further compelling evidence for the importance of a Hox code in specifying segmental identity.

The extensive similarities between MLL/Mll and trx suggest that they are true homologues (Ma et al., 1993). MLL/Mll also contain a motif that is homologous to the cysteine-rich zinc-binding domain of mammalian DNA (cytosine-5) methyltransferases (MTase). It is thought that the function of this motif is to confer the ability to distinguish between methylated and unmethylated DNA. In the absence of its zinc-binding domain, mammalian DNA MTase will promiscuously methylate unmethylated rather than hemimethylated DNA (Kingsley, 1992). This is interesting as it suggests that the function of Mll has evolved to adapt to the presence of methylated sequences in mammalian DNA (see Ma et al., 1993). Vertebrate DNA is typically methylated at a high proportion of C residues occurring in CpG dinucleotides and this is thought to be important for gene inactivation (Meehan et al., 1992). There is no methylation in many invertebrates, however, including Drosophila (Urieli-Shoval et al., 1982). It has been postulated that the vertebrate Hox clusters arose by duplication close to the origin of the vertebrate lineage and that this process also involved the duplication of many other genes and chromosomal regions (Lundin, 1993; Holland & Garcia-Femàndez, 1996). The origin of DNA methylation-based gene repression mechanisms may have permitted the major expansion in the number of functional genes that are present in vertebrates (see Bird, 1995). Therefore, it appears that both the vertebrate Hox clusters and an important mechanism for their regulation have evolved concurrently, further reinforcing the critical interrelationship between structure, function and Hox gene regulation. In this respect it would be interesting to see if a trx homologue without a MTase-like zinc-binding domain exists in more primitive ancestors of vertebrates such as Amphioxus which has a single Hox cluster (Garcia-Femàndez & Holland, 1994).

1.5 Summary

It seems clear that the transcription factors encoded by the Hox/HOM-C genes are key players in a common molecular mechanism for the generation of regional diversity along the embryonic AP axis in many, if not all, animal species (for reviews see McGinnis & Krumlauf, 1992; Kenyon, 1994; Krumlauf, 1994; Lawrence & Morata, 1994). The remarkable structural, organisational and functional conservation among Hox/HOM-C complexes is thought to indicate that a clustered arrangement is crucial for orderly, colinear gene activity and therefore regulation. Much is known about the genetic role played by members of Hox/HOM-C but relatively little is known about how these genes are regulated during embryogenesis.
The characterisation of common factors, engaged in similar aspects of Hox/HOM-C regulation, in Drosophila and mouse implies that at least some regulatory mechanisms may be conserved between the two systems. For example; those involving the interplay between Hox/HOM-C genes via auto and crossregulatory interactions and the involvement of Pc-G and trx-G like factors in maintaining a heritably active or repressed transcriptional state (Paro, 1993; Krumlauf, 1994; Pöppelr et al., 1995). To what extent other regulatory pathways have been conserved, such as those that invoke the initial domains of Hox gene expression in the early embryo, is not known and our knowledge of Hox gene regulation in general is scant. Many broad and specific questions remain to be answered, such as: What are the functional constraints that have maintained the integrity of the Hox/HOM-C complexes throughout evolution? It is possible that this is intimately linked with spatiotemporal, colinear gene expression but the reasons why are obscure (Kenyon, 1994). What are the inductive signals and upstream pathways that lead to the initial activation of Hox gene expression in vertebrates? In Drosophila it is clear that this is achieved by the a hierarchical network of maternally and zygotically derived transcription factors, acting in an acellular environment. Once the Drosophila embryo cellularises, control of HOM-C genes within individual segments is turned over to the segment-polarity genes consisting of the wingless (wg) and hedgehog (hh) signalling systems (Akam, 1987; Perrimon, 1994). Analogous wg/hh systems in vertebrates could be involved in early Hox gene activation (Riddle et al., 1993; Smith, 1994; Roelink, 1996). Other interesting questions, particularly from a mechanistic standpoint, relate to the generation of sharp anterior boundaries of Hox gene expression. For example: What are the molecular mechanisms by which a boundary is set? What are the relative contributions of tissue-specific and general activating, or repressing, factors? Are boundaries within different tissues (e.g. the somites and central nervous system) generated by different mechanisms or by a common mechanism with modifications? Answers to all of these questions will rely, at least in part, on the dissection of the molecular hierarchy that reflects developmental decisions made prior to Hox gene activation.

As previously stated, an important approach towards identifying upstream factors involved in Hox gene regulation is the isolation of cis-regulatory elements, which translate positional information into spatiotemporally restricted transcriptional activity. This has largely been achieved by the employment of reporter gene analyses in transgenic mice. Subsequently, this will allow the characterisation of specific binding-sites and their cognate trans-acting factors, the activity of which represents a previous hierarchical level of patterning information. So far transgenic methods have resulted in the characterisation of a few specific trans-acting factors that are required for particular aspects of Hox gene expression in the developing hindbrain (Sham et al., 1993; Marshall et al., 1994; Studer et al., 1994; Pöppelr et al., 1995; Nonchev et al., 1996). Hopefully, continued research along the same lines will help us to
understand the origin and dissemination of positional information within the mammalian embryo.

1.6 Regulatory organisation of the Hoxb-4 gene

Initial interest in the Hoxb-4 gene as a useful model for studying Hox gene regulation stemmed from the work of Whiting et al. (1991), who showed that a genomic DNA fragment encompassing the gene contained all of the elements required for correct temporal and spatial expression throughout development. During embryogenesis Hoxb-4 expression can be detected at approximately 8.0 dpc from the posterior end of the prospective spinal cord. As time progresses the anterior boundary of neural expression reaches up to a rostral limit at the border between rhombomeres 6 and 7 (r6/7) in the hindbrain. Hoxb-4 is also expressed in derivatives of the neural crest, such as the spinal ganglia and the nodose ganglion of cranial nerve X. In the paraxial mesoderm Hoxb-4 expression is seen in the somites with an anterior boundary between somite 6/7, and subsequently in the prevertebrae with an anterior limit that maps to PV2. This is in accordance with the phenotype of Hoxb-4⁻/⁻ mutant mice which exhibit a partial anterior transformation of the second cervical vertebra (CV2) into the first (CV1) (Ramirez-Solis et al., 1993). Hoxb-4 expression is also found in the mesodermal components of several organs such as the lung, kidney and gut (Graham et al., 1988; Gaunt et al., 1988; Hunt et al., 1991b; Whiting et al., 1991).

Several distinct cis-regulatory elements were defined by Whiting et al. (1991) which are located downstream of the transcription start site; regions A, B and C. Regions A and C are of particular note as they are capable of acting as spatially-specific enhancers and together specify all of the major aspects of endogenous Hoxb-4 expression. The 3kb region A lies at the 3' end of the gene and functions as a neural-specific enhancer, directing expression throughout the neural tube to a sharp anterior limit at the rhombomere r6/7 boundary. The majority of the Hoxb-4 pattern is dictated by the 1.38kb intronic enhancer, region C. This drives expression in the posterior neural tube, derivatives of the neural crest such as the spinal ganglia, and in the mesoderm. Within the paraxial mesoderm it defines the correct anterior somite boundary at the level of somite 6/7 (s6/7). Regions A and C are able to specify these patterns on both the Hoxb-4 and the heterologous hsp68 (heat shock) promoters. Region B, encompassing the Hoxb-4 3'-untranslated region, has not been analysed in detail but seems to contain elements that specify expression in the lung.

The Hoxb-4 promoter (originally defined as 1.2kb of 5'-flanking sequence) is essentially inactive in the embryo, driving a small patch of ectopic expression within the midbrain. Appropriate expression is only seen in the presence of the 3' regulatory elements (Whiting et
We have carried out a detailed analysis of the \textit{Hoxb-4} promoter in cultured cells (Gutman et al., 1994). The architecture of the 5'-flanking DNA is complex and consists of two independent promoters (P1 and P2) situated around two major transcription start sites, located 285 and 207 nucleotides upstream from the ATG translational initiation codon, respectively. The promoter region also contains several \textit{cis}-regulatory elements located upstream and downstream of the transcription start sites. These include three negative regulatory elements, two of which show cell-type specific activity. Positive regulation is achieved by the binding of a novel transcription factor (HoxTF) to the sequence GCCATTGGC, which lies within the 5'-untranslated region. HoxTF is also an essential component of the region C enhancer, as defined in this study.

Analyses in transgenic mice of the \textit{Hoxb-4} genes from the Japanese puffer fish (\textit{Fugu rubripes}) and the chicken, indicate that the overall spatial distribution of \textit{cis}-regulatory elements within these genes seems to be conserved, at least with respect to the anterior domains of neural expression (Aparicio et al., 1995; Morrison et al., 1995). Both the chicken and \textit{Fugu Hoxb-4} loci contain a 3' neural enhancer comparable to mouse \textit{Hoxb-4} region A, and capable of specifying the anterior r6/7 boundary within the hindbrain. However, the regulatory elements or mechanisms specifying \textit{Hoxb-4} mesodermal expression seem to have diverged between the three species. The chicken \textit{Hoxb-4} intron does contain an enhancer that mediates expression in some of the same posterior mesodermal and neural tissues as the mouse intronic enhancer (region C) but, the anterior boundaries of expression are more posterior and the levels of expression lower (Morrison et al., 1995). The intronic sequences from \textit{Fugu} seem to contain no enhancer activity in transgenic mice, despite the presence of a conserved region of sequence homology that is required for both mouse and chicken region C activity (Aparicio et al., 1995; Morrison et al., 1995; see Chapter 4). These disparities may be accounted for either because regulatory elements required for appropriate mesodermal expression of \textit{Hoxb-4} are located elsewhere in the chicken and \textit{Fugu} loci, or because essential mouse transcription factors are incapable of recognising some of the elements within the chicken and \textit{Fugu} introns (Morrison et al., 1995).

### 1.7 Aim of the project

In this study I have begun a detailed analysis of the mouse \textit{Hoxb-4} region C enhancer. The specific aim has been to identify individual \textit{cis}-regulatory binding-sites that are functionally required for the normal activity of the enhancer. Once characterised, these binding-sites could be used to identify and clone the the upstream \textit{trans}-acting factors that interact with them. This approach should enable us to gain insight into the molecular mechanisms underlying spatially and temporally restricted \textit{Hox} gene expression during embryogenesis.
Chapter 3 of this report will describe the structure of region C at the nucleotide level. I will then detail the expression pattern derived from a region C-\textit{lacZ} reporter construct in transgenic mice. Chapters 4 and 5 will describe the identification and characterisation of two independent \textit{cis}-regulatory elements in the \textit{Hoxb-4} intron, using a combination of transgene mutational analyses and biochemical investigations. Chapter 4 focuses the role of HoxTF as a critical component of \textit{Hoxb-4} expression, particularly in the mesoderm. Chapter 5 concentrates on the function of a G-rich binding-site, necessary for appropriate levels of region C enhancer activity. I will illustrate the spatial distribution of regulatory elements within region C and describe a possible mechanism by which these interact, involving the formation of a DNA-loop.
Chapter 2

Materials and Methods

General laboratory safety practices, handling and disposal of radioactive isotopes and other hazardous substances were carried out according to the codes of practice of the National Institute for Medical Research (NIMR). Animal work was performed strictly in accordance with the guidelines set out in the Animals (Scientific Procedures) Act 1986.

Oligonucleotides were supplied by the Sequencing and Synthesis Service Section at the NIMR. Laboratory chemicals were obtained from BDH, Fisons or Sigma and enzymes from Boehringer Mannheim or New England Biolabs (NEB), unless otherwise indicated.

Molecular biological and transgenic techniques used were based on those described by Sambrook et al. (1989) and Hogan et al. (1994), respectively.

2.1 Standard solutions and reagents

All solutions were made up using molecular biology grade reagents, sterile distilled (dH₂O) or deionised (Elga) water, and stored at room temperature unless stated otherwise.

**TE (Tris-EDTA) buffer**
10mM Tris-HCl (pH 8.0), 1mM EDTA (pH 8.0). Sterilise by autoclaving.

**10x dNTPs**
2.5mM dATP, dCTP, dGTP and dTTP. 100mM solutions (Ultrapure, Pharmacia) were diluted to 2.5mM each with TE (pH 7.5) to give a 10x dNTP stock. Store in aliquots at -20°C.

**10x One-Phor-All (OPA) buffer**
100mM Tris-OAc (pH 7.5), 100mM MgOAc, 500mM KOAc. Store in aliquots at -20°C.

**10x ATP**
Stock solution for ligation and kinase reactions. 10mM ATP solution in TE (pH 7.5), store in aliquots at -20°C.
**TBE (Tris-borate EDTA) buffer**
89mM Tris base, 89mM boric acid, 2mM EDTA (pH 8.3). Made as a 5x stock solution and autoclaved. Diluted to 1x as required with the addition of deionised water (Elga). Add ethidium bromide from stock solution to 0.5μg/ml, if required.

**3M NaOAc**
10x salt solution for alcohol precipitation. 3M NaOAc (pH 5.2 with acetic acid).

**TFB I**
30mM KOAc, 100mM RbCl, 10mM CaCl₂, 50mM MnCl₂, 15% (v/v) glycerol (pH 5.8 with acetic acid). Filter sterilise and store at 4°C.

**TFB II**
10mM MOPS-HCl (pH 7.0), 10mM RbCl, 75mM CaCl₂, 15% (v/v) glycerol. Filter sterilise and store at 4°C.

**BHI medium**
3.7% (w/v) brain-heart infusion medium (DIFCO). Autoclave and store at 4°C.

**L-Broth (LB)**
1% (w/v) bacto-tryptone (DIFCO), 0.5% (w/v) bacto-yeast extract (DIFCO) and 0.5% (w/v) NaCl. Sterilise by autoclaving and store at 4°C.

**L-agar**
LB containing 1.5% (w/v) bacto-agar (DIFCO). Autoclaved and cooled to 55°C. Poured as plates (ampicillin added for selective plates) and dried at room temperature for 2-3 days. Store at 4°C.

**Ampicillin 1000x stock**
100mg/ml ampicillin (sodium salt, Sigma) added to bacterial culture media at 100μg/ml. Sterilise by filtration and store in aliquots at -20°C. Media containing ampicillin were stable for 2-3 weeks at 4°C.

**PS I**
25mM Tris-HCl (pH 8.0), 50mM glucose, 10mM EDTA. Store at 4°C.

**PS II**
0.2M NaOH, 1% (w/v) SDS.
PS III
3M KOAc (pH 4.8 with acetic acid). Sterilise by autoclaving.

Phenol: chloroform
1:1 (v/v) ratio of phenol (TE buffered, pH 8.0, Amresco) and chloroform. Store in dark at 4°C.

RNase A stock (250x)
10mg/ml RNase A in 10mM Tris-HCl (pH 7.5), 15mM NaCl. Heated to 100°C for 15 minutes to inactivate any DNase. Store in aliquots at -20°C.

Ethidium bromide stock
10mg/ml stock solution, store at 4°C in dark.

Sequenase reaction buffer (5x)
200mM Tris-HCl (pH 7.5), 100mM MgCl₂, 250mM NaCl. Supplied in kit-form by United States Biochemical (USB). Store at -20°C.

Labelling mix (5x)
dCTP, dGTP and dTTP (7.5µM each). Supplied in kit-form by USB. Store at -20°C.

Termination mixes
80µM each dNTP, 50mM NaCl and 8µM of the particular ddNTP. Supplied in kit-form by USB. Store at -20°C.

Stop solution
95% (v/v) formamide, 20mM EDTA, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol FF. Supplied in kit-form by USB. Store at -20°C.

10x Exonuclease III buffer
660mM Tris-HCl (pH 8.0), 6.6mM MgCl₂. Store at -20°C.

7.4x S1 buffer
300mM KOAc (pH 4.6), 2.5M NaCl, 10mM ZnSO₄, 50% (v/v) glycerol. Store at -20°C.

S1 stop solution
300mM Tris base, 50mM EDTA. Store at -20°C.
10x Klenow buffer
200mM Tris-HCl (pH 8.0), 100mM MgCl₂. Store at -20°C.

Avertin
100% (w/v) 2,2,2-tribromoethanol dissolved in tertiary amyl alcohol. Diluted to 2.5% (w/v) in distilled water. Store avertin solutions at 4°C in the dark.

Microinjection buffer
10mM Tris-HCl (pH 7.4), 0.1mM EDTA. Make up using high quality, endotoxin-free water that has preferably been embryo tested (e.g. W1053, Sigma). Sterilise by passage through a 0.22μM filter. Store in aliquots at -20°C.

PBS (phosphate buffered saline)
138mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄, 1.4mM KH₂PO₄ (pH 7.2). Sterilise by autoclaving.

Embryo fixative
Mirsky’s Fixative (National Diagnostics). Make from 10x concentrate and 10x buffer in distilled water. Store at 4°C for up to 1 month.

Embryo wash solution
0.02% (v/v) Nonidet P40 (NP-40) in PBS.

Embryo stain solution
PBS (pH 7.3) containing 5mM K₂Fe(CN)₆, 5mM K₄Fe(CN)₆, 2mM MgCl₂, 0.02% (v/v) NP-40, 0.5mg/ml X-gal (5-bromo-4-chloro-3-indoly1-β-D-galactopyranoside, Melford Laboratories Ltd.). Store at 4°C in the dark. X-gal can be made up as a 100x stock solution (50mg/ml in N,N-dimethylformamide, store in dark at -20°C). K₃Fe(CN)₆ and K₄Fe(CN)₆ can be made up as 0.5M (100X) stock solutions. Also store at 4°C in the dark.

10x PCR buffer
100mM Tris-HCl (pH 8.3), 500mM KCl, 15mM MgCl₂. Store at -20°C in aliquots.

Micro-extraction buffer
20mM HEPES (pH 7.8), 450mM NaCl, 200mM EDTA, 500mM DTT, 1μg/ml aprotinin, 1μg/ml leupeptin, 100μg/ml PMSF and 1μg/ml pepstatin A. Make fresh from stock solutions.
2x Binding buffer
100mM Tris-HCl (pH 7.9), 12mM MgCl2, 0.4mM EDTA, 2mM DTT, 30% (v/v) glycerol.
Store in aliquots at -20°C.

2.2 DNA manipulations

2.2.1 Digestion and preparation of DNA for subcloning

Unless otherwise stated, restriction endonuclease digestions were performed under the conditions recommended by the manufacturer. Typically, digests were performed for 1-2 hours with 0.1-1μg of DNA and 0.5-10 units of enzyme per 10μl of reaction volume.

To minimise non-recombinant background in cloning procedures, digested vector DNA was treated with calf intestinal alkaline phosphatase (CIAP, Boehringer Mannheim) prior to ligation. After digestion, 0.1 units of CIAP were added to the reaction followed by incubation at 37°C for 30 minutes. The phosphatase was then inactivated at 85°C for 30 minutes.

For blunt-ended ligations 5'-overhanging termini were filled in by the addition of 1x dNTPs (250μM) and 1 unit of Klenow enzyme (Boehringer Mannheim) to the digested DNA, followed by incubation at room temperature for 20 minutes. Klenow was inactivated at 70°C for 20 minutes. 3'-overhanging termini were removed by the addition of 1x dNTPs and 1 unit of T4 DNA polymerase (Boehringer Mannheim), followed by incubation at 37°C for 30 minutes. Inactivation conditions were the same.

Oligonucleotides used for cloning steps, EMSA or PCR were supplied fully deprotected in ammonia solution. They were recovered by ethanol precipitation, pelleted, washed with 70% (v/v) ethanol and resuspended in TE or water. The ethanol wash step was necessary to remove any ammonium ions that could inhibit subsequent kinase reactions. Concentrations were determined by measuring the spectrophotometric absorbance at 260nm (assuming for oligonucleotides an OD of 1 absorbence unit = 33μg/ml).

Prior to ligation 0.5μg of each complementary oligonucleotide were phosphorylated by the addition of 10 units of T4 polynucleotide kinase (PNK, New England Biolabs) to a 10μl reaction volume containing 1x One-Phor-All buffer (OPA, Pharmacia) and 1mM ATP. After incubation at 37°C for 30 minutes, PNK was inactivated and the oligonucleotides annealed by heating to 70°C for 15 minutes, followed by gradual cooling to room temperature.
2.2.2 Agarose gel electrophoresis and DNA fragment isolation

DNA fragments generated by restriction endonuclease digestion were separated by electrophoresis through 0.7-1.5% (w/v) agarose in 1x TBE containing 0.5μg/ml ethidium bromide. The products were then visualised under ultra-violet light and photographed using a Polaroid Land Camera fitted with a Wratten filter (No. 25, Kodak) and Polaroid 667 black and white film.

Specific restriction fragments for subcloning were fractionated by agarose gel electrophoresis. A well was cut out of the gel, in front of the appropriate band, using a sharp blade. The well was then filled with molten 1% (v/v) low-melting-point (LMP) agarose (Seaplaque, Flowgen), allowed to set at 4°C, and electrophoresis was continued until the DNA had migrated into the LMP agarose. The band was cut out and melted at 65°C. Agarose was digested with 2 units of β-Agarase I (NEB) at 40°C for 1-2 hours with the supplied buffer. Undigested carbohydrates were pelleted by adding NaOAc (pH 5.2) to 0.3M, chilling on ice for 15 minutes and centrifugation (14 000rpm for 15 minutes). DNA was recovered by isopropanol precipitation.

2.2.3 DNA ligations

All reactions were performed in 1x OPA buffer in the presence of 1mM ATP and 1 unit of T4 DNA ligase (Boehringer Mannheim) in a volume of 10-20μl (overnight at 16°C or 2 hours at room temperature). Ligations contained 1-5ng/μl of linearised, dephosphorylated, vector DNA and a 2-50 fold molar excess of insert DNA.

2.2.4 Preparation of competent host bacteria

Most of the transformation-competent bacteria used in this study were prepared by C. Halai of the Division of Eukaryotic Molecular Genetics, NIMR. *Escherichia coli* (E. coli) bacteria of the DH5-α or NM522 strains were cultured in 500ml of L-broth (LB) at 37°C with aeration (170rpm, gyratory shaker), until they had reached an optical density of 0.5 at 600nm. The cells were rapidly chilled and harvested by centrifugation at 0°C prior to resuspension in 100ml of TFB I at 0°C. They were then pelleted again and resuspended in 10ml TFB II at 0°C. Aliquots of the resuspended cells were flash-frozen on dry ice and stored at -80°C.
2.2.5 Transformation of competent cells

DNA constructs were introduced into bacterial cells using the standard protocol. In general, 1-10ng of supercoiled plasmid DNA or 10-20μl of a ligation reaction were used to transform 100-200μl of cells. Frozen stocks of competent *E. coli* were hand-thawed and incubated with the DNA for 15 minutes on ice, followed by a 90 second heat shock at 42°C. Cells were allowed to recover in 500μl of warm LB at 37°C for 40 minutes. They were then pelleted, resuspended in 200μl of LB and spread onto L-agar plates containing 100μg/ml ampicillin to select for resistant transformants. Colonies that formed after overnight incubation at 37°C were picked and grown for further analyses.

2.2.6 Isolation of plasmid DNA

Plasmid clones were propagated in the *E.coli* host strains DH5-α or NM522, grown in BHI medium (DIFCO) containing 100μg/ml ampicillin. Clones were screened by restriction analysis of DNA isolated by a modified alkaline lysis 'miniprep' method (Sambrook *et al.*, 1989).

Overnight bacterial cultures (1.5 ml) were pelleted in a microfuge. The pellets were resuspended in 200μl of PS I, lysed with 200μl of PS II and cleared with 200μl of PS III. Samples were then extracted by shaking with 500 μl of phenol:chloroform and centrifuged at 14,000rpm for 2 minutes. The aqueous phase (500 μl) was transferred to a fresh tube, precipitated with isopropanol and pelleted at 14,000rpm for 15 minutes. DNA pellets were washed with 70% (v/v) ethanol, air dried and resuspended in 30μl of TE containing 40μg/ml RNase A. 2-5μl was used for restriction analysis or 10-20μl for DNA sequencing.

For more extensive manipulations, DNA samples were prepared from 250ml overnight cultures by the alkaline lysis method and further purified by equilibrium centrifugation in a CsCl-ethidium bromide gradient (see Sambrook *et al.*, 1989). Some of these bulk preparations of plasmid DNA were made by C. Halai at the NIMR.

Cultured bacteria were harvested by centrifugation (8000rpm, 20 minutes, 20°C) in a Sorvall RC-5B centrifuge, fitted with a GSA rotor, then resuspended in 20 ml of PS I. Cells were lysed by gentle inversion with 40 ml of PS II and the lysate treated by the addition of 30 ml of PS III. After centrifugation (GSA rotor, 10,000rpm, 15 minutes, 4°C) the supernatant was filtered through 2 layers of tissue (Kleenex medical wipes, Kimberley-Clarke) and the plasmid DNA precipitated with 0.6 volumes (54 ml) of isopropanol. The previous centrifugation step was repeated and the DNA pellet rinsed with 70% (v/v) ethanol, dried and resuspended in 5 ml of TE. 4.75 ml of this solution was used to dissolve 5.1g of CsCl in a
plastic universal (Sterilin) and 0.25ml of ethidium bromide (10mg/ml) were added. The tube was left to stand at room temperature for 5 minutes to allow a precipitate to form, which was then pelleted in a Sorvall RC-6000B centrifuge (1500rpm, 5 minutes, 20°C). The supernatant was transferred to a polyallomer ultracentrifuge tube (Beckman Quick-Seal #342412) which was heat-sealed. Supercoiled plasmid DNA was banded by ultracentrifugation in a Beckman VTi 80 rotor for 4 hours (70 000rpm, 20°C). Banded DNA was removed from the centrifuge tube with a syringe and extracted with 2-5 volumes of water-saturated butanol to remove the ethidium bromide. CsCl was removed by dialysis against TE buffer or by diluting to 2.5 volumes with distilled water, followed by alcohol precipitation at room temperature.

DNA concentrations were determined by spectrophotometric absorbance at 260nm (assuming for double stranded DNA an OD of 1 absorbence unit = 50µg/ml), or estimated according to ethidium bromide fluorescence when electrophoresed against DNA samples of known size and concentration. Samples were stored as solutions at -20°C (unless used for transfections where to reduce nicking they were stored at 4°C) or as 20% (v/v) glycerol stocks of bacterial cultures at -80°C.

2.2.7 DNA sequencing

Nucleotide sequences were determined by the dideoxynucleotide chain termination method (Sanger et al., 1977) using the Sequenase Version 2.0 Sequencing Kit (United States Biochemical).

2-8µg of plasmid template DNA were denatured by the addition of NaOH to 0.4M (5 minutes at room temp.) and neutralised by the addition of an equal volume of 1M Tris-HCl (pH 7.5). The sample was diluted to 200µl with water and the DNA recovered by ethanol precipitation. After washing with 70% (v/v) ethanol, the denatured DNA pellet was resuspended in a total volume of 10µl with 2µl 5x Sequenase reaction buffer and a equimolar amount of oligonucleotide primer. After brief vortexing 1µl 0.1M DTT, 2µl 5x labelling mix, 0.5µl [α35S]-dATP (370MBq/ml, >15TBq/mmol, Amersham) and 2µl (3.25 units) of diluted Sequenase T7 DNA polymerase were added. For sequencing close to the primer (30 bases) 5-fold less labelling mix was used. Labelling was performed for 2-5 minutes at room temperature before 3.5µl of the labelling reaction were added to 2.5µl of each of the four chain termination mixes (G, A, T, or C) containing the appropriate dideoxynucleotide. The termination reactions were incubated at 37°C for 5 minutes followed by the addition of 4µl of stop solution. The samples were then denatured at 80°C for 2 minutes, immediately prior to resolution on a buffer-gradient polyacrylamide gel (Biggin et al., 1983; Littlebury, 1993). Gels were fixed in a solution of 10% (v/v) acetic acid and 10% (v/v) ethanol for 15 minutes,
dried at 80°C under vacuum and exposed to Kodak XAR or Biomax film at room temperature.

2.2.8 Generation of nested deletions

For the deletional analysis of region C reporter constructs in transgenic mice a series of 5' and 3' nested deletions (Henikoff, 1984; Sambrook et al., 1989) were made as follows:

10μg of pHLC-II DNA were digested to completion in a 100μl reaction volume with either SalI and PstI (5' deletions) or BglII and SphI (3' deletions), as verified by ethidium bromide-agarose gel electrophoresis of a 300ng (3μl) sample. The resulting DNA was extracted once with an equal volume of phenol:chloroform and precipitated with ethanol. After microcentrifugation at 13 000rpm for 10 minutes the DNA pellet was washed with 70% (v/v) ethanol, dried, resuspended in 60μl of 1x exonuclease III buffer and equilibrated at 37°C in a water bath. Meanwhile, 7.5μl of S1 reaction mixture containing 1x S1 buffer and 2.25 units of S1 nuclease (Boehringer Mannheim) were aliquotted into 25, 0.5ml microcentrifuge tubes on ice. 2.5μl of the DNA solution were transferred to the first tube (time 0) and 300 units of exonuclease III (Boehringer Mannheim), equivalent to 150u/pmol of 5'-overhanging termini, were added to the remaining DNA and mixed. The reaction was immediately returned to 37°C and 2.5μl samples were removed at 30 second intervals for 23 subsequent time-points. Samples were then incubated at room temperature for 30 minutes prior to the addition of 1μl of S1 stop solution, followed by incubation at 70°C for 10 minutes to inactivate the S1. To determine the extent of digestion 2μl (40-80ng) aliquots of the various time points were analysed by electrophoresis through a 0.7% (w/v) ethidium bromide-agarose gel. The points containing suitable deletion ranges (up to 1.5kb) were selected. Remaining 3'-overhanging termini were removed by addition of 1μl of Klenow mix, containing 1x Klenow buffer and 3-5 units of Klenow fragment (Boehringer Mannheim), followed by incubation at 37°C for 3 minutes. DNA strands were repaired by the addition of 1μl of 10x dNTPs (2.5mM each) for a further 5 minutes. Samples were ligated at room temperature for 1 hour by adding 40μl of ligase mix (1x OPA, 1mM ATP, 0.2 units T4 DNA ligase) and transformed into bacteria. Resulting clones were screened by restriction endonuclease digestion and suitably sized clones sequenced to determine the exact extent of deletion.
2.3 Production and analysis of transgenic mice

2.3.1 Transgenic reporter constructs

All constructs for microinjection were liberated with \textit{NotI} unless otherwise indicated.

\textbf{WT:} The 3.9kb \textit{SalI-BglII} fragment from p610Za-BII (kindly provided by R. Allemann) containing the \textit{hsp68-lacZ-SV40} poly(A) reporter gene (Whiting \textit{et al.}, 1991) was end filled and cloned into the \textit{EcoRV} site of pPolyIII-I (Lathe \textit{et al.}, 1987) to give pHl-I. Two complementary oligonucleotides;

\[\text{ND+ (S'CTTCGAAAGATCTGGTCACCGTCGACATTATCTGCAGGCGGCCGCG)}\]
\[\text{ND- (S'-GATCCGCGGCCGGCGCCTGCAGATTAATGTCGACGGTGACCAGATCCTTCGAAGCATG)}\]

were annealed and cloned between the \textit{SphI} and \textit{BglII} sites of pHl-I, destroying the latter site by a \textit{BamHI/BglII} fusion and modifying the polylinker to give pHl-II. The 1.38kb \textit{SalI-BglII} fragment of \textit{Hoxb-4} was subcloned into pHl-II to give the basic wild-type reporter construct which was used for subsequent manipulations (pHLC-II).

\(\Delta 514-616\): pHLC-II was linearised with \textit{MunI} then partially digested with \textit{SfiI}. The \textit{MunI} and \textit{SfiI} sites were filled in and chewed back, respectively, to yield blunt-ended termini by the addition of T4 PNK and dNTPs, followed by religation to give pHLC-\(\Delta 514-606\). This construct carried an internal deletion of bp 514-606 of region C.

\(\Delta 558-598\): Two complementary oligonucleotides;

\(\Delta \text{CB+ (S'-AATTGTCCCGCTATAAATCTCCATTGCCCAGAGATTACGTCCTCCTGGCCTAGT)}\)
\(\Delta \text{CB- (S'-AGGCCAGAGACCGTAATCTCGGCAATGGGAGATTTATAGCGGGGAC)}\)

were annealed and cloned into the \textit{MunI} and \textit{SfiI} (partial) sites of region C to generate pHLC-\(\Delta 558-598\). This construct carried an internal deletion of bp 558-598 of region C.

\textbf{HoxTF/YY1 mutants:} In order to enable a simplified cloning procedure for the \textit{MunI-SfiI} area of region C without the need for partial \textit{SfiI} digestion, the \textit{SfiI} polylinker site of pHLC-II was knocked out by digestion with \textit{XhoI}, end filling, and religation to give pHLC-III.

Mutations in the HoxTF/YY1 binding-sites were generated by cloning two pairs of complementary partially overlapping oligonucleotides into the \textit{MunI-SfiI} sites of pHLC-III. The right hand complementary pair;

\(\text{MSMB (S'-TCTCCTGTTCAGACGACATAATTACATCGCCCATAAATTATGGCCTAGT)}\)
\(\text{MSMD (S'-AGGCCATAAAAATTTATGGGGCGATGTAAATTATGGCCTCGA)}\)

were common to each construct. The left hand complementary pair carried the desired mutated bases (underlined):
MH/Y:
MH/Y+ (5'-AATTGTCCCCGCTATAAACTCGCAGG GCCAGAGATTTACGG ) and
MH/Y- (5'-AAACAGGAGACCCTAATCTCTGTC CCTGCGAGTTTATAGCGGGGAC)

MHTF1:
MHTF1+ (5'-AATTGTCCCCGCTATAAACTCGCCATT TTAG GAGATTTACGG) and
MHTF1- (5'-AAACAGGAGACCCTAATCTCTCCT AAATGG ATGCCC TTATAGCGGGGAC)

MHTF2:
MHTF2+ (5'-AATTGTCCCCGCTATAAACTCGCCATT TTAG GAGATTTACGG) and
MHTF2- (5'-AAACAGGAGACCCTAATCTCTCCT AAATGG ATGCCC TTATAGCGGGGAC)

MYY1:
MYY1+ (5'-AATTGTCCCCGCTATAAACTCGCCATT TTAG GAGATTTACGG ) and
MYY1- (5'-AAACAGGAGACCCTAATCTCTCCT AAATGG ATGCCC TTATAGCGGGGAC)

1XHTF: A single copy of the HoxTF binding-element (bp 510-557 of region C) was cloned
in a 5' to 3' direction upstream of the hspa promoter to generate pHL-1XHTF. Two
complementary oligonucleotides;
HTF+(5'-AGCTTAATTGTCCCCGCTATAAACTCGCCAGTGCCAGAGATTTACGGTCTCCT) and
HTF- (5'-AGCTAGGAGACCGTAAATCTCTCGCACTGGCGAGTTTATAGCGGGGACAATTA),
carrying a point mutation in the YY1 binding-site (underlined), were cloned into the HindIII
site of p610ZAI (Whiting et al., 1991).

3XHTF: 3 head to tail copies of the above oligonucleotides were cloned similarly, upstream
of the hspa promoter of p610ZAI to generate pHL-3XHTF. Both 1XHTF and 3XHTF were
digested with HindIII and Asp718 for microinjection.

$\Delta$475: This construct (pHLC-$\Delta$475) was isolated as one of a number of nested deletions
generated from the 5' end of region C as described in section 2.2.8, deleting the first 474bp.

$\Delta$514: pHLC-II was digested with MunI and NotI for microinjection, deleting the first 513bp
of region C.

$\Delta$819: pHLC-II was digested with Asp718 and BglII, end filled, and religated to yield
pHLC-475-818, deleting the 3' 563 bp of region C.
Δ475MG: Two complementary oligonucleotides were cloned into the SalI-MunI sites of pHLC-III; MUTG+ (5'-TCGACTCCGAGGCAATGTTITCGAGGGAGCGAGGGGAAGCC) and MUTG- (5'-AATTGCTTCCCCCTCGCTCCCTCCAAACATGCCTCGGAG). The construct is identical to Δ475 but carries a null mutation in the Granville binding site (underlined).

G+HTF: The complementary oligonucleotides; G+HTF+ (5'-AATTGTCCCCGCTATAAACCCTCGACAGAGTTTACGA) and G+HTF- (5'-GATCTCGTAAATCTCTGGCAATGGCGAGTTTATAGCGGCGGAC) were cloned into the MunI-BglII sites of pHLC-Δ475 to generate pHL-G+HTF. This construct carries bp 475-550 of region C encompassing the Granville and HoxTF/YY1 binding-sites.

2.3.2 Preparation of DNA for microinjection

Linear fragments of DNA for microinjection were excised by digestion of the vector (10-50 μg) with the appropriate restriction enzymes (as indicated in 2.3.1). Fragments were then recovered and purified from agarose gels as described in Chapter 2.2.2. Prior to alcohol precipitation the DNA was further purified by extraction, once with an equal volume of phenol:chloroform and twice with chloroform. Each of these extraction steps were performed with care so as not to carry over any of the organic phase. After recovery of the DNA from solution by alcohol precipitation and centrifugation, the DNA pellet was washed thoroughly, several times, with 70% (v/v) ethanol and dried under vacuum for 5-10 minutes. The DNA was resuspended in microinjection buffer and passed through a 0.2μM filter-centrifugation unit (Spinex, Costar). DNA concentrations were estimated by comparative ethidium bromide fluorescence after electrophoresis against DNA samples of known size and concentration. Transgenic mice were produced by microinjecting the DNA (1-5g/ml in microinjection buffer) into the pronuclei of fertilised one-cell mouse embryos (Hogan et al., 1994).

2.3.3 Transgenic procedures

Throughout these experiments (CBA x C57BL10) F1 mice were used as embryo donors, stud and vasectomised males, pseudopregnant females, and breeding females. Animals were supplied by the SPF Breeding Unit at the NIMR and were maintained in a constant environment on a 12 hour light-dark cycle (05.00-21.00hrs light period). All regulated procedures performed, were licensed under the Animals (Scientific Procedures) Act 1986.

Twelve female mice (4-week old) were superovulated (Hogan et al., 1994) by injection with
5IU/0.1ml of pregnant mare's serum (PMS), between 15.00-16.00hrs. A second intraperitoneal injection of 5IU/0.1ml of human chorionic gonadotrophin (hCG) was administered 46 hours later. Hormones were obtained from Intervet Laboratories as Folligon (PMS) and Choluron (hCG), resuspended at 50IU/ml in sterile PBS and stored at -20°C in aliquots. Following hCG injection each female was placed in a cage with a stud-male, then removed the following morning and checked for a vaginal plug.

Oviducts from plugged females were removed and the eggs were dissected out into ambient temperature M2 medium. Cumulus cells were removed by transferring the embryos into M2 containing 300μg/ml hyaluronidase for 2-5 minutes. After washing thoroughly in M2 to remove any debris, embryos were transferred to KSOM medium microdrop-cultures under mineral oil (Sigma #M8140), equilibrated in an incubator at 37°C with 5% (v/v) CO₂. M2 (see Hogan et al., 1994) and KSOM (Bhatnagar et al., 1995) were obtained from Speciality Media Inc., Lavallette, New Jersey. Media were stored at -20°C for up to two months. Once thawed, they were filter sterilised and stored in aliquots at 4°C for up to 1 week.

DNA constructs were microinjected into either of the pronuclei of a one-cell embryo, contained in a droplet of M2 under oil. Generally, embryos were microinjected in batches of 20-30, then washed through equilibrated KSOM and returned to the incubator. The microinjection setup was mounted on a vibration-free table and consisted of a Nikon Diaphot-TMD inverted microscope fitted with Nomarski optics, combined with Leitz-E micromanipulators.

Embryos which survived the microinjection procedure were reimplanted into the oviducts of pseudopregnant plugged females (Hogan et al., 1994). Recipient mothers were prepared by mating 6-8 week old females in natural oestrus with vasectomised males and were supplied by the SPF Breeding Unit at the NIMR. 10-15 embryos were transferred into each oviduct of a female, anaesthetised by intraperitoneal injection with 0.35ml of 2.5% (w/v) avertin. This time-point was arbitrarily defined as 0.5 days post coitum (dpc).

The embryos that developed were either allowed to proceed to term for the production of transgenic lines, or harvested transiently. Transgenic embryos carrying lacZ reporter constructs were analysed for β-galactosidase activity as described below.

2.3.4 β-Galactosidase assay

Embryos were dissected free from uteri, decidua and extraembryonic membranes in PBS. Where necessary placental tissues were retained for transgenic identification as described
in Chapter 2.3.5. Embryos were transferred to glass liquid-scintillation vials, rinsed in PBS to remove serum and cellular debris, then fixed for 15-90 minutes (depending on size) in Mirsky's Fixative (National Diagnostics) at room temperature. After washing for 30 minutes at room temperature in embryo wash solution, the fixed embryos were stained in the dark at 37°C in embryo stain solution for 16-20 hours. Embryos were then washed in PBS to remove any stain solution, fixed again as described above and stored at 4°C in 70% (v/v) ethanol (after dehydration through a series of 30% (v/v) and 50% (v/v) ethanol).

2.3.5 Transgenic detection by PCR

Transgenic embryos or animals were identified by PCR amplification of a transgene specific fragment (450bp), spanning the junction of the lacZ gene and SV40 poly(A) signal, using the primers; LZ3 (5’-GCGACTTCCAGTCAACATC) complementary to the minus strand of lacZ, and STB (5’-GATGAGTGGACGAAAACAC) complementary of the plus strand of the SV40 poly(A) sequence. As a qualitative internal control for the PCR reaction the primer pair MGP1 (5’-CCAAGTTGGTGCAAGAC) and MGP2 (5’-CTCTCTGCTTTAAGGAGTCAG) were used to amplify a 172bp fragment from the endogenous myogenin gene (Yee & Rigby, 1993).

Tail biopsies (0.5cm) were taken from mice at 3 weeks of age, or placental tissue was retained following the dissection of F₀ embryos. The samples were placed in 300µl of 1x PCR buffer containing 0.1% (v/v) Tween-20 and heated to 100°C for 5-10 minutes. Proteinase K was added to a final concentration of 100µg/ml and following overnight digestion at 55°C, cellular debris was pelleted by microcentrifugation. PCRs were performed in a total reaction volume of 20µl, overlaid with mineral oil in a 96-well plate (Thermowell, Costar), using a PTC-100 thermal cycler (MJ Research, Genetics Research Instrumentation). Each reaction contained 1x PCR buffer (1.5mM MgCl₂), 1x dNTPs (250µM), 50ng of each primer, 1.5u of Taq DNA polymerase (Amplitaq, Perkin Elmer Cetus) and 1µl of digested tissue sample. The amplification reaction conditions used were; 94°C/3 mins., 28 cycles of 94°C/30 secs., 55°C/30 secs., 72°C/30 secs., followed by 72°C for 2 mins. PCR products were resolved by ethidium bromide-agarose gel electrophoresis.

2.3.6 Histological studies

Embryos were wax-sectioned by W. Hatton of the Sectioning and Histology Service at the NIMR. Blue staining embryos, stored in 70% (v/v) ethanol, were dehydrated through 85%
(v/v), 95% (v/v) and 2 changes of absolute ethanol. Dehydration steps were performed at room temperature for 30-60 minutes according to the size of the specimen. After clearing (2 x 30 minute changes of Histoclear, National Diagnostics) at room temperature, embryos were impregnated with paraffin wax (Fibrowax, BDH) by incubating at 60°C with 1:1 (v/v) Histoclear:wax, followed by 3 changes of wax for 30-60 minutes each. Specimens were transferred to disposable plastic moulds containing fresh molten wax, oriented as required and allowed to cool. Wax blocks were stored desiccated at 4°C. Sections were cut at 6μm thickness and collected on TESPA coated slides (Rentrop et al., 1986), dewaxed with changes of Histoclear (2 x 5 minutes) and serially passed through 3 changes of absolute, 1 x 95% (v/v), 1 x 70% (v/v) ethanol and 1 x distilled water (1 minute each). Sections were then counterstained with eosin (0.5% (w/v) in 25% (v/v) ethanol) for 10-30 seconds. Excess eosin was removed by 2 x 10 second rinses in distilled water and samples were dehydrated through 70% (v/v), 95% (v/v) and 2 x absolute ethanol (15 seconds each). Sections were then cleared through 4 changes of Histoclear (2-3 minutes each) and mounted under glass coverslips with DPX mountant (BDH #360294H).

2.4 DNA electrophoretic-mobility shift assay (EMSA)

2.4.1 Probe labelling and purification

Oligonucleotides, prepared as described in 2.2.1, were end-labelled with 32P, annealed, and used as probes in EMSA and UV cross-linking experiments. Complementary oligonucleotides (2pmol each) were labelled in a 10μl reaction containing 1x OPA buffer, 6μl [γ32P]-ATP (370MBq/ml, >185TBq/mmol. Redivue, Amersham) and 10 units of T4 PNK (NEB) for 30 minutes at 37°C. The reaction was heated to 70°C for 20 minutes to inactivate the enzyme and the oligonucleotides annealed at room temperature for 1 hour to overnight.

The DNA probe was separated from single-stranded DNA and unincorporated radiolabelled nucleotides on a native polyacrylamide gel. Gels (8-12% dependent on probe size) were prepared from a stock solution of 30% (w/v) acrylamide/1.034% (w/v) bisacrylamide in 1x TBE (Easigel, Scotlab) and run at 4-8W for 3-5 hours. After electrophoresis the gel was exposed to Kodak XAR film for 10 minutes and the resulting autoradiograph used as a template for the excision of the probe-containing gel slice. The probe was eluted overnight in 200μl of distilled water at room temperature, transferred to a fresh tube and stored at -20°C. The specific activity of freshly labelled probe was approximately 1000cpm/fmol and was usable for up to 2 weeks. The recovery of the labelled probe was generally 50-75% by this method, as judged by comparing the cpm emitted by the eluted versus non-eluted probe.
Specific competitor DNA was prepared by annealing 1μg of each unlabelled, complementary oligonucleotide in a 10μl volume, followed by dilution to an appropriate concentration.

2.4.2 Preparation of protein extracts

Some of the extracts from cell-lines and embryos were kindly provided by Dr Alejandro Gutman of the Division of Eukaryotic Molecular Genetics (NIMR). The method used was that described by Scholer et al. (1986) for the preparation of whole-cell micro-extracts except that the sonication step was replaced with three cycles of freeze-thawing. Briefly, approximately 10^7 cells were pelleted and resuspended in 1x micro-extraction buffer then transferred to an ethanol-dry ice bath until frozen. Thawing was achieved by incubation at 30°C. Samples were microcentrifuged (14 000rpm for 10 minutes at 4°C) and the supernatant aliquotted, assayed for protein concentration, and stored at -80°C.

Protein concentration was determined using the Bradford micro-assay method (Bradford, 1976), by comparison against known amounts of standard protein (BSA, Sigma). Samples were diluted to a final volume of 800μl in distilled water, 200μl Bio-Rad Protein Assay reagent was added, and the contents mixed by inversion. The samples were allowed to stand at room temperature for 10 minutes before reading the absorbence at 595nm.

2.4.3 EMSA

DNA binding reactions were set up in a 20μl volume and contained; 1x binding buffer, 1μg of sonicated salmon sperm DNA, specific competitor (where indicated) and 10-20μg of whole-cell extract. After incubation on ice for 15 minutes, 1μl (10fmol) of probe was added and the incubation continued for an additional 10 minutes at 30°C. Samples were kept on ice before being loaded on to a 5% (w/v) polyacrylamide gel. Electrophoresis was performed using BRL vertical gel electrophoresis apparatus (fitted with a recirculating pump) in 0.25x TBE (150 volts for 2 hours at 4°C). The gel was dried under vacuum at 80°C and exposed to Kodak XAR or Biomax film with an intensifying screen at -80°C.

2.4.4 UV cross-linking

UV cross-linking experiments (Chodosh et al., 1986) were performed by Dr Alejandro Gutman of the Division of Eukaryotic Molecular Genetics (NIMR). The probe (HoxC-WT) was prepared as described in 2.4.1. The complementary oligonucleotides used (made by the
Sequencing and Synthesis Service Section, NIMR) consisted of the region C sequences from bp 523-549 and contained the modified base 5-bromo-deoxyuridine (BrdU) instead of deoxythymidine (T). A preparative DNA-binding reaction was made (5-fold scale up) in a standard 1.5ml microfuge tube (Scotlab). Following the probe incubation step, bound polypeptides were crosslinked to the probe by exposure to UV light (305nm wavelength, 5cm distance) then resolved on an 8% (w/v) SDS-polyacrilamide gel. The gel was dried under vacuum at 80°C and exposed to Kodak XAR film with an intensifying screen at -80°C. The apparent molecular weights of the detected polypeptide species were estimated by comparison against known standards (Biorad).

2.5 Transfection experiments

Transfection experiments and CAT (chloramphenicol acetyltransferase) assays were performed by Dr Alejandro Gutman. The Hoxb-4 minimal promoter construct (MP) has been described previously (Gutman et al., 1994) and consists of the -32 to +21 sequences cloned into the vector pBLCAT3 (Luckow & Schütz, 1989). WT 1X includes one copy an oligonucleotide that encompasses the sequences from bp 510-557 of region C and was made by cloning the following complementary oligonucleotides; WT+ (5'-AGCTTAATTGTCCCCGCTATAAACTCGCCATTGCCAGAGATTTACGGTCTCCT), and WT- (5'-AGCTAGGAGACCGTAAATCTCTGGCAATGGCGAGTTTATAGCGGGGACAATTA), into the HindIII site of MP. WT 3X contains three head to tail copies of the same oligonucleotide and MH/Y, MHTF2 and MYY1 are the same as WT 1X except that they carry the mutations as indicated in Figure 4.6.

The mouse neuroblastoma cell-line Neuro2a (ATCC# CCL 131) was cultured and maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 4mM L-glutamine, 100u/ml penicillin, 10mg/ml streptomycin (GIBCO) and 10% (v/v) fetal calf serum (Imperial Laboratories Europe). Cells were passaged every 2-3 days (seeded at a density of approximately $5 \times 10^4$ cells/ml) and maintained in an incubator at 37°C with 5% (v/v) CO$_2$. Sub-confluent 9cm diameter plates were transfected by the standard calcium phosphate method (Sambrook et al., 1989) with 16pg of a reporter recombinant and 4 μg of the internal control plasmid RSV-βgal (Gorman et al., 1983). Cells were incubated with the calcium phosphate precipitate in complete medium for 20 hours, then washed with PBS and grown for an additional 28 hours in complete medium. Duplicate experimental points were determined in three independent transfections. β-galactosidase and CAT activities were measured as described previously (Sleigh, 1986; Sambrook et al., 1989).
Chapter 3

Characterisation of a *Hoxb-4* region C/heat shock promoter reporter construct

Transgenic experiments using an *E. Coli* ß-galactosidase reporter gene (*lacZ*) have demonstrated that the expression patterns of some *Hox* genes can be fully or partially recreated when a genomic DNA fragment is isolated from a *Hox* complex. Further analysis has lead to the identification of both general and tissue-specific regulatory elements. In a few cases trans-acting factors that interact with specific elements have been defined (Sham *et al.*, 1993; Marshall *et al.*, 1994; Studer *et al.*, 1994; Pöpperl *et al.*, 1995; Nonchev *et al.*, 1996). Despite these efforts our understanding of *Hox* gene regulation remains limited. The mapping of distinct cis-regulatory units is an essential step towards understanding the functional organisation of the *Hox* complex. A transgenic approach using *lacZ* reporter genes under the control of genomic sequences provides a valid *in vivo* assay for pinpointing such elements. This will ultimately enable us to identify the upstream factors that are responsible for establishing and maintaining the unique characteristics of *Hox* gene expression.

The *Hoxb-4* gene provides a useful model for the study of *Hox* gene regulation. Whiting *et al.* (1991) have shown that a genomic DNA fragment encompassing the *Hoxb-4* gene contains all of the elements required for correct temporal and spatial expression during development. Several distinct regulatory regions have been defined which contribute to various aspects of this pattern. Of particular note are two enhancers which specify practically all of the features of endogenous *Hoxb-4* expression. Region A functions as a neural-specific enhancer directing expression throughout the neural tube to a sharp anterior limit at the rhombomere 6/7 (r6/7) boundary. The majority of the *Hoxb-4* pattern, however, is dictated by region C. This enhancer drives expression in the posterior neural tube, derivatives of the neural crest such as the spinal ganglia and in the mesoderm. Within the paraxial mesoderm region C defines the correct anterior somite boundary at the level of somite 6/7 (s6/7), and is able to specify this pattern on both the *Hoxb-4* and the heterologous *hsp68* (heat shock) promoters.

Region C comprises the intron and parts of the flanking exons of *Hoxb-4*. The aim of this project has been to further analyse this enhancer in detail. By locating smaller cis-regulatory components within region C it should be possible to identify the specific trans-acting factors that bind to them. This will enable us to gain insight into the molecular mechanisms underlying spatially and temporally restricted *Hox* gene expression during embryogenesis.
Figure 3.1: Genomic organisation of the mouse *Hoxb-4* gene

The genomic organisation of the *Hoxb-4* locus spanning 6.2kb is illustrated. The minimum sequences required for correct *Hoxb-4* expression in transgenic mice are contained within a 6.1kb *NcoI* fragment (previously misquoted as 7.4kb) and the relative locations of the previously defined regulatory elements; regions A, B and C, are shown above the locus in green (Whiting *et al.*, 1991). Exonic sequences are denoted by boxes and the polyadenylation signal (poly A) is marked. The first transcriptional start site (P1) is indicated by a broken arrow (Gutman *et al.*, 1994). The vertical blue and black bars represent the translational initiation (ATG) and termination codons, respectively. The red shaded region denotes the homeobox within exon 2. Also shown is the relative position of a conserved block of sequence homology (CB1) which lies within the intronic sequences of the region C enhancer (light blue oval).

Abbreviations of restriction enzyme sites: A (*Asp718*), B (*BglII*), H (*HindIII*), M (*MunI*), N (*NcoI*), P (*PstI*), R (*Rsal*), S (*Sall*), Sf (*SfiI*).
Figure 3.2: Sequence of the 1.38kb region C fragment

The nucleotide sequence of the 1.38kb (*SalI-BglII*) *Hoxb-4* region C enhancer is shown. The position and recognition sequences of the major internal restriction enzyme sites are labelled and overlined. Numbering is with respect to the first nucleotide of the *SalI* recognition site (GTCGAC), marked with an asterisk (+1). The amino acid sequence of the *Hoxb-4* protein corresponding to exons 1 and 2 is displayed above the nucleotide sequence and the intron/exon boundaries are marked with arrowheads. The nucleotide and amino acid sequences of the homeobox and homeodomain, respectively, are underlined in exon 2 (bp 1245-1381). Also shown are a number of features that are referred to later in this text. Namely; the location of the HoxTF/YY1 and Granville binding-sites (boxed); the sequences corresponding to the conserved block of sequence homology (CB1) within the intron underlined in bold (bp 511-606); the 5' extent of the the Δ475 construct marked with a broken arrow.
Val Asp Pro Lys Phe Pro Pro Cys Glu Glu Tyr Ser Gin Ser Asp Tyr Leu Pro Ser Asp
GTCGACCCCAAGTTCCCTCCGTGCGAGGAGTATTCACAGAGGCTATTACCCTACCCAGGCCAC

His Ser Pro Gly Tyr Tyr Ala Gly Gly Gin Arg Arg Glu Ser Gly Phe Gin Pro Glu Ala
CACCTCCGGGGTGCTACAGCCGGCGGCAAGGAGGCTATTACCCTACCCAGGCCAC

Ala Phe Gly Arg Arg Ala Pro Cys Thr Val Gin Gin Tyr Ala Ala Cys Arg Asp Pro Gly
GCTTTTGGCCTCCGGCGGCGGCCTACATGGCAGCCTGACGGCTATTACCCTACCCAGGCCAC

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro
GCCCGCCACCTCGGCCCCGCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
3.1 DNA sequence of Hoxb-4 region C

The genomic organisation of the Hoxb-4 locus and the localisation of previously identified regulatory regions within it are shown in Figure 3.1. Region C is defined as a 1.38kb SauI-BglII fragment incorporating the Hoxb-4 intron (Whiting et al., 1991). The nucleotide sequence of the intron has been determined previously (Dr. R. Allemann, personal communication) and verified by the sequencing of various plasmid clones during the course of this study, as described in the Materials and Methods Chapter (2.2.7). The sequence of region C showing the conceptual translation of the coding exonic sequences, exon/intron boundaries, and other relevant sequence data is presented in Figure 3.2. The Hoxb-4 intron is 800bp in length and throughout this study the numbering of nucleotides within region C is with respect to the first base of the SauI recognition site (GTCGAG) in exon I. This corresponds to +321 nucleotides, relative to the transcription start site of the Hoxb-4 gene (Gutman et al., 1994).

3.2 Expression of a region C-lacZ reporter gene in transgenic mice

The introductory aim of this project was to assemble a region C reporter construct which could be employed in transgenic mice. This would be used to further define the developmental expression pattern dictated by region C. Having determined a baseline, smaller cis-regulatory sequences could be delimited by comparing the expression pattern derived from mutant constructs to that of the wild-type.

The cloning steps used to generate the wild-type region C construct (WT) are described in the Materials and Methods Chapter (2.3.1). WT comprises region C cloned upstream of a reporter construct consisting of the hsp68 promoter driving lacZ, with the SV40 termination and polyadenylation sequences attached to the 3' end (Whiting et al., 1991). The hsp68 promoter was employed for several reasons. Firstly, region C is able to confer spatial-specificity on the hsp68 promoter and impose an expression pattern on it very similar to that seen with the Hoxb-4 promoter. Secondly, it provides relatively high levels of reporter gene expression compared to those seen with the Hoxb-4 promoter. It was envisaged that this would be advantageous for the analysis of mutant constructs where low levels of expression might be expected. Thirdly, the heterologous promoter provides a context in which to study region C enhancer activity in isolation from other regulatory elements that are known to be present within the Hoxb-4 promoter (Gutman et al., 1994).

WT was microinjected into the pronuclei of one-cell mouse embryos as described in the Materials and Methods Chapter (2.3.3). Transgenic mice generated in this way were initially
examined by transient analysis of F₀ embryos at 10.5 days post coitum (dpc). This was the developmental time point previously described for the expression of an equivalent construct (Whiting et al., 1991). Of the six transgenic embryos obtained, all four that expressed the transgene did so in a manner consistent with that previously described (data not shown). These data indicated that WT was functional and capable of giving appropriate region C-directed expression. In addition to the transient analysis a transgenic line was established which was mated to wild-type animals to obtain embryos at various developmental time-points. This enabled the expression pattern derived from this region C reporter to be studied throughout embryogenesis, which has not been previously done. The results of these analyses are shown in Figure 3.3.

Transgene expression driven by region C (WT) was initially activated in the late pre-somite stage embryo at 7-7.5 dpc. Staining due to β-galactosidase activity was detectable within the ectoderm, mesoderm and endoderm posterior to the neural groove, and also at the base of the allantois (Figure 3.3a). At 8.5 dpc (10-12 somite stage) strong β-galactosidase staining was visible within the posterior neuroectoderm and neural tube, then more weakly up to an anterior limit at the presumptive spinal cord/hindbrain boundary (Figure 3.3b). This is more posterior than the normal boundary of endogenous Hoxb-4 expression within the neuroectoderm which extends to the level of the r6/7 boundary (Graham et al., 1988; Whiting et al., 1991). Mesodermal staining was widespread and located in the somitic, presomitic and lateral mesoderm. Somitic staining was visible in the most posterior, newly formed, somite up to an anterior limit at the level of s6/7. In a analogous manner to the neural expression pattern the most anterior somites were much more weakly stained.

The expression pattern observed at early stages was maintained throughout later stages of development. Between 9.5 and 12.5 dpc strong staining could be clearly seen in the neuroectoderm and neural crest derivatives, such as the spinal ganglia, and in the somitic and lateral mesoderm, posteriorly from the forelimb (Figure 3.3c-g). Anterior limits of expression in the spinal cord and somites remained weak, however, becoming stronger more posteriorly. This is somewhat different from the pattern observed when region C is used to drive expression from the Hoxb-4 promoter. In this case the levels of expression are strongest at the anterior boundaries, becoming weaker more posteriorly. By 15.5 dpc, β-galactosidase detectability is impaired due to the developing skin acting as a barrier to the fixation and staining processes. At this stage, staining could be seen in the developing mammary glands (data not shown) and in the primordia of sensory hair follicles of the face (Figure 3.3h), located above the eye (Kaufman, 1992). This particular aspect of follicular expression is different from that seen with the full length Hoxb-4-lacZ transgene previously described (Whiting et al., 1991). In this latter case β-galactosidase activity was noted only in the dermal placodes of the skin with an obvious axial limit at the cervical level.

84
Figure 3.3: Developmental time-course of region C/hsp68-lacZ reporter gene expression

A time-course of transgenic embryos carrying the mouse Hoxb-4 region C enhancer within construct WT is illustrated. A schematic of the WT reporter is shown below. The hsp68-lacZ-SV40pA reporter gene is not to scale. Major restriction enzyme sites are marked and the legend to region C follows that described in Figure 3.1 and Figure 4.1. Exp. denotes the total number of positively stained lines (1) and transient embryos (4) obtained that showed a consistent pattern of expression. Tg. denotes the total number of transgenic founders and embryos.

(a) Lateral view of a 7.5 dpc transgenic embryo showing staining in the ectoderm, mesoderm and endoderm posterior to the caudal tip of the neural groove (arrowhead). The anterior end of the embryo is to the left (a) and the posterior end to the right (p).  (b) Dorsal view of an 8.5 dpc embryo, with the anterior end at the top of the picture. Strong staining can be seen in the posterior neuroectoderm and neural tube, extending to an anterior limit at the presumptive spinal cord/hind brain boundary. Note that the levels of neural tube staining at the anterior boundary are weak. Widespread staining is also visible within the lateral, presomitic and somitic mesoderm. Somitic staining is similarly weak towards its anterior margin at the boundary between somites 6 and 7 (s6 and s7). Only feeble staining is visible in s7, indicated by an open arrowhead. The boundaries between somites are indicated by arrows.  (c) Lateral view of a 9.5 dpc transgenic embryo, by which stage the anterior somite boundary is clearly visible at the border between s6 and s7 (arrow).  (d) and (e) Lateral and dorsal views of a 10.5 dpc transgenic embryo. An anterior limit of strong staining within the neural tube is visible (>), progressing much more weakly up to the spinal cord/hindbrain boundary marked by an arrow.  (f) A transverse section through the forelimb bud (lb) of a similar 10.5 dpc embryo (10X magnification), where staining can be seen in the somite (s) and in the neural tube (nt).  (g) Lateral view of a 12.5 dpc embryo.  (h) Lateral view of the head of a 15.5 dpc embryo. Staining can be seen in the primordia of the follicles of vibrissae around the snout (hf) and in the prominent tactile hair follicles (>).
3.3 Summary and Discussion

The results have so far described preliminary studies into transcriptional regulation by the Hoxb-4 enhancer, region C. Initial studies have involved the production of a region C/hsp68-lacZ reporter gene and have documented its activity throughout embryonic development in transgenic mice. Spatial and temporal patterns of expression given by this WT transgene mirror those described by Whiting et al. (1991) for an analogous region C/hsp68 construct. My experimental data suggest that the region C/hsp68 transgene is activated in a manner temporally consistent with the endogenous Hoxb-4 gene and the region C/Hoxb-4 promoter construct previously described (Whiting et al., 1991). With the WT construct β-galactosidase activity is first detected in pre-somite stage embryos at 7-7.5 dpc. This is slightly earlier than that described for the full length Hoxb-4 construct and expression is visible in all three germ-layers; ectoderm, endoderm and mesoderm. The endoderm is not a site of expression of the endogenous Hoxb-4 gene, nor of any previously documented Hoxb-4 transgenes. This anomalous result may reflect an increase in β-galactosidase detectability due to the strength of the hsp68 promoter, relative to that of Hoxb-4. It could also be due to endogenous β-galactosidase activity present within the early endoderm, as reported by Vogels et al. (1993) when studying the regulation of the Hoxb-7 gene. The major aspects of the spatial expression pattern and tissue distribution derived from WT are consistent with those of the region C/Hoxb-4 promoter construct. In particular, the boundaries of expression within the central nervous system and the somitic mesoderm are maintained. However, some differences in the relative levels of expression are evident. Most notably, anterior boundaries of expression in the mesoderm and neuroectoderm are weak with the WT construct and expression becomes stronger more posteriorly. This may reflect functional interactions between regulatory elements within region C and the Hoxb-4 promoter that are necessary to achieve correct levels of spatial expression.
Chapter 4

HoxTF: a critical factor for \textit{Hoxb-4} regulation

4.1 Sequence comparison identifies a conserved region within the intron

Having defined a suitable system for the assay of region C transcriptional activity \textit{in vivo}, the next step of my research was to adopt a logical strategy to facilitate the localisation of \textit{cis}-regulatory elements within this enhancer. Transgenic analysis is labour and time intensive and prohibits the use of a systematic mutagenesis approach, such as a linker-scanning, which would require a large number of mutant constructs to be tested. DNase-hypersensitive site mapping provides another rationale for identifying regions of the genome that are involved in \textit{cis}-regulation. However, this technique is most suitable for the analysis of homogeneous cellular populations such as cultured cell-lines. The application of this method to the study of protein-DNA interactions within region C in an embryonic context, where multiple modes of tissue-specific regulation are likely, was deemed to be too technically demanding and may have provided results that would have been difficult to interpret.

Sequence comparison, on the other hand, provides a powerful method for identifying functionally conserved sequences at both the nucleic acid and protein level. The structural, functional and regulatory conservation between the \textit{Hox/HOM} complexes has been well documented (Akam, 1989; Duboule & Dollé, 1989; McGinnis & Krumlauf, 1992; Kappen & Ruddle, 1993), and the existence of homologous and paralogous genes within the vertebrate \textit{Hox} clusters that possess very similar patterns of expression, and presumably regulatory mechanisms, makes comparative sequence analysis a rational approach to help define conserved \textit{cis}-regulatory elements. Despite the relative simplicity of this strategy it suffers from an inherent drawback which is akin to looking for 'a needle in a haystack'. Noncoding DNA constitutes approximately 95% of the mammalian genome and is largely composed of repetitive or 'junk' sequences that have no known regulatory function. Consequently, searching for conserved regions that represent \textit{bona fide} \textit{cis}-regulatory elements, that may amount to only a few to tens of nucleotides, is fraught with problems of detection. In an attempt to overcome this problem it is possible to compare sequences from widely divergent vertebrate species, the assumption being that at least some of the most fundamental regulatory regions may have been conserved throughout evolution. Thus, interspecies sequence comparison between mouse and chicken noncoding regions enabled Gérard \textit{et al.} (1993) to quickly localise regulatory domains within the \textit{Hoxd-11} gene. To develop this strategy one step further, three way evolutionary sequence comparisons should provide an
additional level of stringency by filtering out some areas of sequence conservation that occur between any two species as a matter of chance, or nonfunctional similarity, rather than because of a conserved functional role.

The Japanese puffer fish \textit{(Fugu rubripes)} belongs to the family \textit{Tetradontidae}, the members of which are remarkable since, amongst vertebrates, they possess highly compacted genomes. The haploid \textit{Fugu} genome is approximately 400Mb in length, some 7.5 times smaller than those of mammals (approximately 3000Mb). This compaction is due to a reduced abundance of noncoding DNA such as repetitive sequences, that constitute less than 10\% of the \textit{Fugu} genome, with a general reduction in intron size (reviewed in Elgar \textit{et al}., 1996). This property makes \textit{Fugu} an attractive proposition for sequence comparison between noncoding regions of DNA, since similarities due to functional conservation should be more evident. Another advantage of \textit{Fugu} is that teleosts represent the most evolutionarily distant taxonomic group from mammals within the vertebrate lineage, with a separation time of some 430 million years. As most distant relatives, sequence identities between mouse and \textit{Fugu} should be maximally biased towards functional conservation. In collaboration with Dr. S. Aparicio and Dr. S. Brenner (Molecular Genetics Unit, University of Cambridge, Cambridge, U.K.), we were interested in using \textit{Fugu Hoxb-4} sequences for evolutionary comparison with those of the mouse.

Intronic sequences from Paralogous Group 4 (PG-4) \textit{Hox} genes from mouse, chicken and \textit{Fugu} were subjected to DNA alignments. The mouse \textit{Hoxa-4} sequence was obtained from the EMBL DNA database under accession number X66861 (Galliot \textit{et al}., 1989). The intron of mouse \textit{Hoxd-4} was subcloned from a genomic clone (a gift of Dr. M. Featherstone, McGill University, Montréal, Canada) as an \textit{Accl-BglIII} fragment into pBluescript KS+ (Stratagene) and sequenced using T3/T7 primers as described in the Materials and Methods Chapter (2.2.7). Chicken \textit{Hoxa-4}, \textit{b-4} and \textit{d-4} sequences were kindly provided by Dr. A. Kuroiwa (Nagoya University, Nagoya, Japan). The \textit{Fugu Hoxb-4} gene was cloned and sequenced by Dr. S. Aparicio as part of the aforementioned collaboration (Aparicio \textit{et al}., 1995). Initial pairwise sequence alignments were made between mouse \textit{Hoxb-4} and the other sequences using the Macvector Pustell DNA Matrix program (Macvector Sequence Analysis Programs, IBI-Kodak). The parameters used for the alignments were; window size = 30, minimum \% score = 65\%, hash value = 6.

This approach identified a single conserved block of homology (CB1) present within the introns of all PG-4 genes examined (Aparicio \textit{et al}., 1995; Morrison \textit{et al}., 1995). The relative positions of CB1 within the mouse \textit{Hoxb-4} locus and region C are shown in Figures 3.1 and 3.2, respectively. The conserved sequences were imported into a multiple alignment.
Figure 4.1: Sequence conservation between the introns of PG-4 Hox genes

A schematic of the Hoxb-4 region C enhancer is presented above showing the relative position of the conserved block of nucleotide sequence homology (CB1) within the intron. CB1 is represented as a pale blue oval. The vertical black bar within CB1 represents a transcription factor binding-site for HoxTF/YY1 that is discussed in Chapter 4. Exonic sequences are denoted by boxes, the pale yellow box represents the homeobox in exon 2. Major restriction enzyme sites are marked as described in Figure 3.1. A=Asp718; B=BglII; M=MunI; S=SalI; Sf=SfiI.

An alignment of sequences conserved between the introns of Paralogous Group 4 Hox genes from mouse (M), chicken (C) and Fugu rubripes (F) is shown below. Bases that are identical to the 96bp CB1 sequence of mouse Hoxb-4 are highlighted. The numbering of bases within region C is shown above the alignment. The position of a transcription factor binding-site for HoxTF/YY1 (GCCATTGCC) lying between bp 531-539 of region C is denoted by a red line. The bases deleted in constructs Δ514-606 and Δ558-598, discussed in Chapter 4.2, are over lined. The position of highly similar sequence motifs (I-IV) conserved amongst different paralogues, are underlined at the bottom.
program (MegaAlign, DNASTAR Inc.) and aligned manually to give an optimal match to the mouse Hoxb-4 sequence. The results of these alignments are shown in Figure 4.1. The greatest degree of identity is seen between the Hoxb-4 sequences (mouse-chicken = 74% over 96bp; mouse-Fugu = 59%; average = 67%). Hoxa-4 sequences show an intermediate identity (average = 49%) and Hoxd-4 sequences the least (average = 41%), over the same region. All sequences, however, share common identity over the most highly conserved 30bp region (70-90%), corresponding to bp 571-601 of region C. Several motifs of sequence conservation that exhibit the same relative spacing are evident from this alignment. These motifs are common to either all, or some, homologous groups (I-IV in Figure 4.1). Motif I is found in all Hoxb-4 and a-4 sequences but not in Hoxd-4. Motif II is confined to Hoxb-4, although nucleotides similar to it (CTGTNNNNNAG) are located in the Hoxa-4 sequences but displaced 3' by 5bp (bp 561-572 relative to region C in Figure 4.1). Motifs III and IV are common to all PG-4 genes examined. Motif III contains an ATT A core sequence (bp 575-578) corresponding to a potential homeodomain binding-site (Gehring et al., 1994). Motif IV contains sequences which closely resemble the consensus binding-sites for both the Drosophila hunchback and the Drosophila and vertebrate caudal-related group of proteins (Zuo et al., 1991; Suh et al., 1994). Wu & Wolgemuth (1993) have previously described a functional in vitro binding-site for the Hoxa-4 protein within the Motif IV sequences of the mouse Hoxa-4 intron. However, the TAAT core of this site is not conserved amongst any of the other PG-4 genes examined (bp 594-598 of region C in Figure 4.1). In addition to the sequence identity to mouse Hoxb-4 over CB1, other homologous genes show strong similarities over this same region. Mouse and chicken Hoxa-4 are highly conserved with approximately 85% sequence identity. Mouse and chicken Hoxd-4 are less conserved sharing approximately 50% identity (data not shown).

4.2 Sequences within CB1 are essential for correct region C activity

I wished to investigate whether the conserved sequences within the Hoxb-4 intron are required for cis-regulation. To test this, two mutant reporter constructs were made that carried internal deletions of CB1, the extents of which can be seen in Figure 4.1. The cloning steps used to generate the two mutant reporters (A558-598 and A514-606) are described in the Materials and Methods Chapter (2.3.1). The constructs were assayed in transgenic mice that were analysed as transient F0 embryos between 10.5 and 11.5 dpc. The results are shown in Figure 4.2.

A deletion of 41bp that removed the CB1 sequences corresponding to Motifs II, III and IV, had no significant effect on the expression pattern of the transgene (construct A558-598). Six positively stained embryos were obtained with this construct that showed a consistent pattern
Figure 4.2: CB1 sequences are essential for correct region C activity

The effects of two different deletions of CB1 sequences on the expression pattern of the region C/lacZ transgene are illustrated. Construct Δ558-598 carries a deletion of 41bp removing the most conserved sequences at the 3'-end of CB1. Construct Δ514-606 carries a larger deletion of 97bp encompassing the whole of CB1 (see Figure 4.1). A schematic of both constructs and of the parental region C construct (WT) is shown below. The deleted regions are represented by broken lines, otherwise the legend to the reporters is as described in Figure 3.3. Exp. denotes the total number of positively stained embryos obtained that showed a consistent pattern of expression. Tg. denotes the total number of transgenic embryos.

(a) Shows a lateral view of a 11.0 dpc transgenic embryo carrying the WT region C construct for comparison.

(b) Lateral and (c) dorsal views of an 11.5 dpc transgenic embryo carrying construct Δ558-598. The pattern of β-galactosidase staining obtained is comparable to that of the WT construct.

(d) Lateral and (e) dorsal views of an 11.5 dpc transgenic embryo carrying construct Δ514-606. No expression is detectable within the somitic mesoderm or PNS. Staining is restricted to the ventral region of the neural tube (V). The strong expression around the eye is unique to this particular embryo and was not seen in any of the other five expressers. The neural tube expression does not extend as posteriorly as that of the WT region C construct but does extend more rostrally than the anterior limit at the spinal cord/hindbrain boundary. (f) Shows a transverse section taken at the axial level of the forelimb (40X magnification). Staining within the ventral neural tube (V) can be seen in both the ventricular zone (VZ) and the post-mitotic cells of the marginal zone (MZ). The staining spreads laterally through the commissural neuron pathway (dashed line) and down into the ventral motor horn and may mark a subset of presumptive motor neurons. No staining is visible within the dorsal root ganglion (d).
of β-galactosidase activity that was comparable to that of WT at 10.5 and 11.5 dpc (Figure 4.2a-c). However, a larger deletion of 97bp that removed all of the CB1 sequences had a much more drastic effect on the activity of the transgene (construct Δ514-606). This mutation abolished all expression within the somites, peripheral nervous system (PNS) and the majority of the central nervous system (CNS). A consistent low level of expression within the ventral neural tube was visible in all five of the positively stained embryos obtained, at 12.5 dpc (Figure 4.2d-e). This staining, present in both the ventricular and marginal zones of the spinal cord (Figure 4.2f), corresponds to the portion of the neural tube where motor neurons are born and differentiate and may mark a subset of these cells.

These data show that one or more positive cis-regulatory elements are located within the CB1 sequences of Hoxb-4, and that they are required for efficient region C-directed expression in the mesoderm, neural tube and PNS. Furthermore, the results suggest that the elements lie within the sequences deleted in construct Δ514-606 but not those deleted in construct Δ558-598, i.e. between bp 514-557 and/or bp 599-606 of region C (see Figure 4.1).

4.3 Identification of overlapping binding-sites for HoxTF and YY1 in CB1

To characterise the trans-acting factor(s) that activates transcription via CB1, I performed DNA electrophoretic-mobility shift assays (EMSA). Overlapping oligonucleotide probes were designed that spanned the bp 514-606 sequences, i.e those deleted in the transgenic analyses. The possibility remained that the element(s) may be located at the extremities of the sequences covered by the bp 514-606 deletion and not within it. To allow for this, probes were designed to extend beyond these sequences at the 5' and 3' ends. Oligonucleotide probes were labelled with [γ32P]-ATP, annealed, and purified by native polyacrylamide gel electrophoresis. Whole-cell extracts from 10.5 dpc embryos and the mouse neuroblastoma cell-line Neuro2a were used as sources of proteins. Neuro2a cells were used as they are known to contain both HoxTF and YY1 binding activities (Gutman et al., 1994). EMSA experiments were carried out as described in the Materials and Methods Chapter (2.4) and the products resolved on 5% (w/v) polyacrylamide gels. The results of these experiments are summarised in Figure 4.3.

Specific DNA binding was confined to a probe covering bp 508-567 of region C. Under the experimental conditions used, no specific binding could be seen to the sequences deleted in construct Δ558-598, which include Motifs II, III and IV and are those most highly conserved between paralogues (probes 537-567, 562-596 and 582-616). Further subdivisions of the bp 508-567 sequences failed to show any binding (probes 508-524, 514-536 and 537-567), suggesting that sequences present in both 514-536 and 537-567 are required. This hypothesis
was confirmed by the probe 524-550 (HoxC-WT) which behaved as 508-567 and detected two major retarded bands (see lane 1 in Figure 4.5A).

To further delimit the binding-site(s) within HoxC-WT I systematically mutated the 529-544 sequences. Four probes were designed which carried contiguous 4bp mutations in the HoxC-WT sequences (HoxC-M1-4). The introduced mutations were of the least conservative variety, i.e. A to C, C to A, G to T and T to G transversions. Mutated probes were assayed for their ability to compete for the HoxC-WT binding activity in EMSA experiments, the results of which are summarised in Figure 4.3B (data not shown). The HoxC-M1-4 oligonucleotides were also labelled and used as probes in EMSA experiments. HoxC-M1 and HoxC-M2 showed no specific binding, whilst HoxC-M3 and HoxC-M4 behaved as HoxC-WT (compare lanes 1 and 6-9 in Figure 4.5A). These experiments revealed the presence of two overlapping binding-sites localised to bp 529-536 of region C, where the sequence TCGCCATT is partially or totally required for the formation of both retarded bands.

Concurrent with this line of investigation, we identified a similar binding-site within the promoter of the Hoxb-4 gene that binds the transcription factors HoxTF (homeobox gene transcription factor) and YY1. The Hoxb-4 promoter, originally defined by Whiting et al. (1991) as 1.2 kb of 5'-flanking DNA, is essentially inactive in transgenic embryos and can direct consistent lacZ expression only within a confined region of the midbrain, where the endogenous gene is not expressed. The promoter must contain regulatory elements that are capable of interacting with downstream enhancers, such as region C, and translating this regulatory information into proper levels of spatial gene activity. This point is exemplified by my results which show that on the heterologous hsp68 promoter, region C is unable to impose the appropriate spatial distribution of transcriptional activity especially with regard to sharp anterior boundaries of expression (Chapter 3). Dr Alejandro Gutman of the Division of Eukaryotic Molecular Genetics (NIMR), in collaboration with myself, has investigated the transcriptional regulation of the Hoxb-4 promoter in cultured cells, where it is active (Gutman et al., 1994). A summary of the findings of those experiments is presented in Figure 4.4.
A summary of the results of EMSA experiments used to localise the binding of factors within CB1 is illustrated.

(A) Shows the Hoxb-4 intron sequences, spanning CB1, encompassing bp 508-616 of region C (numbered above, upper strand only). Double-stranded oligonucleotide probes were derived from this sequence for use in EMSA experiments. The sequences deleted in the transgenic constructs Δ558-598 and Δ514-606 (Figure 4.2) are underlined in black.

(B) Shows the boundaries of double-stranded probes used in EMSA experiments. Black oblongs denote probes that detected no specific binding. Red oblongs denote probes that formed two specific complexes. The numbers shown above define the limits of each probe and correspond to the region C sequence (Figure 3.2). Specific binding of the two bands was found to be confined to the wild-type, 27bp, 524-550 probe (HoxC-WT). Part of the sequence of HoxC-WT (bp 529-544) is shown within it. These sequences were systematically mutated in probes HoxC-M1, M2, M3 and M4. The mutated bases are underlined. HoxC-M1 and M2 did not bind but M3 and M4 did, defining the sequences TCGCCATT (bp 529-536) as being partially or totally required for the formation of both bands (see Figure 4.5A, lanes 6-9).
A

CCAATTGTCCCCGCTATAAACCTGCCATTGCCAGAGATTACGGTCTCCTGTTTTCAGAG

\( \Delta 514-606 \)

\( \Delta 558-598 \)

CCAATAATTACATCGCCCATTAATTTTATGGCCTAGTGCGGCTGCGC

B

HoxC-WT  TCGCCATTGCCAGAGA
HoxC-M1  CATCCATTGCCAGAGA
HoxC-M2  TCGCAAGGCCAGAGA
HoxC-M3  TCGCCATTACGAGA
HoxC-M4  TCGCCATTGCCATCTC
Figure 4.4: Model of the mouse *Hoxb-4* promoter

A schematic representation of the *Hoxb-4* promoter is illustrated (after Gutman *et al.*, 1994). The annotations the accompany the diagram follow the key that is shown below.

P1 and P2 indicate the positions of the two independent promoters, containing transcriptional initiators, and their associated transcription start sites. The numbering shown is with respect to the P1 transcription start site (+1). The relative positions of regulatory elements a, b, c and d are indicated. Elements a, b and d act in a negative fashion, with elements a and d exhibiting cell-type-specific activity. The precise location of element b is uncertain; it could lie between +21 and +47 or between +85 and +113 (as is shown). Positive regulation is conferred by element c that contains a HoxTF/YY1 binding-site, essential for efficient *Hoxb-4* expression. The HoxTF/YY1 binding-site is shown in its approximate location and the asterisks above the nucleotides CCA denote bases that are essential for binding.
\( +1 \)  
\( +79 \)  
\( \text{gccatggc} \)

**Diagram Key:**
- \( \text{PI} \) = promoter 1
- \( \text{PII} \) = promoter 2
- \( i \) = initiators
- \( \text{a,b,c,d} \) = regulatory elements
- \( \text{E Z 3} \) = negative regulatory elements
- \( \text{ATG} \) = initiator codon
- \( \square \) = cell-specific elements
element, located within the 5'-untranslated region of *Hoxb-4* between positions +143 and +169 relative to PI, and it can activate transcription in a number of different promoter contexts. Transcriptional activation by element c is mediated via the binding of HoxTF, putatively a novel transcription factor, to the sequence GCCATTGGC that is essential for the efficient expression of *Hoxb-4* promoter constructs in cultured cells. Overlapping the HoxTF binding-site is an independent binding-site for the zinc finger protein Yin-Yang 1 (YY1; also known as delta, F-ACT-1, NF-E1). YY1 is a versatile transcription factor that can act as an activator, repressor, or initiator of transcription. It is able to bind to sites that frequently contain a core motif of CCAT and is important for the regulation of a wide range of unrelated viral and cellular genes (Hariharan *et al.*, 1991; Lee *et al.*, 1991, 1994; Park & Atchison, 1991; Shi *et al.*, 1991; Flanagan *et al.*, 1992). However, in our cell culture transfection experiments, we were unable to demonstrate that the binding of YY1 to element c played any role in transcriptional regulation of the *Hoxb-4* promoter (Gutman *et al.*, 1994).

To identify other potential binding-sites for HoxTF, I conducted a search of the EMBL DNA sequence database using the GCGFIND application, belonging to the University of Wisconsin Genetics Computer Group (UWGCG) suite of programs (Devereux *et al.*, 1984), on a Sun Sparcstation computer. The sequence of the HoxTF/YY1 binding-site from the *Hoxb-4* promoter (GCCATTGGC) was used as a search-string with up to two mismatches allowed. As this would produce numerous 'hits', many of them within protein coding sequences, I reduced the field of sequence entries searched to only those from vertebrate species that contain keywords pertaining to noncoding regions of the genome (intron, promoter, 5'-untranslated region and 3'-untranslated region). This step effectively enriches the sequences searched for noncoding DNA with a potential regulatory function. Two HoxTF binding-sites that were identified in this way were found to lie within upstream regions of the murine myogenic regulatory factor genes, *myogenin* (accession number M95800) and *MyoD1* (accession number X61655) (Zingg *et al.*, 1991; Yee & Rigby, 1993; Gutman *et al.*, 1994). We were immediately interested in these binding-sites since both *myogenin* and *MyoD1*, like *Hoxb-4*, are expressed within the paraxial mesoderm during embryogenesis, where they play a key role in the differentiation of the skeletal musculature.

To investigate the binding characteristics of the *Hoxb-4* intron binding-site (HoxC-WT), I performed additional EMSA experiments. Initially, the binding activities of HoxC-WT and a probe corresponding to the HoxTF/YY1 site from the *Hoxb-4* promoter (HoxP-WT; Gutman *et al.*, 1994) were compared. Both probes detected two bands with similar mobilities (Figure 4.5A, lanes 1 and 2), strongly suggesting that the site in the intron also binds HoxTF and YY1. The binding of the YY1-like band was approximately equivalent for both probes. However, HoxC-WT bound HoxTF with 5- to 10-fold lower affinity than HoxP-WT. A probe comprising a similar HoxTF site from *myogenin* detects only one band that co-migrates
Figure 4.5: *In vitro* analysis of the HoxTF/YY1 binding-site in CB1

(A) The oligonucleotide probes used are described below and are indicated above each lane. Probes were labelled with $[^{32}\text{P}]$-ATP as described in the Materials and Methods Chapter (2.4). Whole-cell extracts from the mouse neuroblastoma cell-line Neuro2a were used as a source of proteins (15µg per reaction), as this cell-line is known to contain both the HoxTF binding activity and YY1 (Gutman *et al.*, 1994). The products of the binding reactions were resolved on a 5% (w/v) polyacrylamide gel. The two major retarded bands corresponding to either HoxTF or YY1 interactions are indicated. FREE indicates the unbound probe.

HoxC-WT encompasses the bp 524-550 sequences of region C and HoxC-M1 to M4 represent mutated versions of this probe used to define the binding-site within it (Figure 4.3). HoxP-WT encompass sequences +143 to +169 from the *Hoxb-4* 5'-untranslated region (Gutman *et al.*, 1994). Myo-WT encompass sequences +9 to +35 from the mouse *myogenin* promoter (Yee & Rigby, 1993). Myo-M is like Myo-WT but carries the mutation CCA to AAC in bases +17 to +19. MD1-WT encompass sequences -628 to -602 of the mouse *MyoD1* gene (Zingg *et al.*, 1991).

(B) Indicates the sequences of the HoxTF binding motifs in the *Hoxb-4* (intron and promoter), *myogenin* and *MyoD1* genes. A consensus binding-site derived from these four sources is presented below.
A

Probe

HoxC-WT  HoxP-WT  Myo-WT  Myo-M  MD1-WT  HoxC-M1  HoxC-M2  HoxC-M3  HoxC-M4

HoxTF →

YY1 →

FREE →

1 2 3 4 5 6 7 8 9

B

<table>
<thead>
<tr>
<th>Location</th>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoxb-4 intron</td>
<td>GCCATGGGC</td>
</tr>
<tr>
<td>Hoxb-4 promoter</td>
<td>GCCATGGGC</td>
</tr>
<tr>
<td>Myogenin</td>
<td>GCCAGTGGC</td>
</tr>
<tr>
<td>MyoDl</td>
<td>GCCATCCC</td>
</tr>
<tr>
<td>Consensus</td>
<td>GCCATGGCC</td>
</tr>
</tbody>
</table>

G

CC
Figure 4.6: *In vitro* analysis of mutations specific for HoxTF and YY1

(A) The HoxC-WT probe was labelled with $[^{32}\text{P}]$-ATP as described in the Materials and Methods Chapter (2.4). Whole-cell extracts from the mouse neuroblastoma cell-line Neuro2a were used as a source of proteins (15µg per reaction). Specific competitor was added at 100-fold molar excess and is indicated above each lane. The products of the binding reactions were resolved on a 5% (w/v) polyacrylamide gel. The two major retarded bands corresponding to either HoxTF or YY1 interactions are indicated. FREE indicates the unbound probe.

(B) Shows the sequence of the upper strand of the wild-type and mutated region C oligonucleotides (HoxC-WT; -MH/Y; -YY-1; -HTF1 and -HTF2), together with their binding capabilities (+ = binding; - = no binding; +/- = partial binding). The sequence of the wild-type HoxTF site is boxed in HoxC-WT and mutated bases are circled in the other sequences. HoxP-WT and HoxP-M represent wild-type and mutant probes from the *Hoxb-4* promoter. Myo-WT and Myo-M represent wild-type and mutant probes from the *myogenin* promoter (see Figure 4.5).
B

<table>
<thead>
<tr>
<th>Probe</th>
<th>HoxC-WT</th>
<th>HoxC-MYY1</th>
<th>HoxC-MHTF2</th>
<th>HoxC-MH/Y</th>
<th>HoxP-WT</th>
<th>HoxP-M</th>
<th>Myo-WT</th>
<th>Myo-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxC-WT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxC-MYY1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxC-MHTF2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxC-MH/Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxP-WT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxP-M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myo-WT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myo-M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**A**

- None
- HoxC-WT
- HoxC-MYY1
- HoxC-MHTF2
- HoxC-MH/Y
- HoxP-WT
- HoxP-M
- Myo-WT
- Myo-M

**B**

<table>
<thead>
<tr>
<th>Probe</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxC-WT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxC-MYY1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxC-MHTF2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxC-MH/Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxP-WT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoxP-M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myo-WT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myo-M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
with HoxTF, whereas that from *MyoD1* also binds YY1 though quite weakly compared to HoxC-WT (see Figure 4.5A, lanes 3 and 5).

In order to confirm the identities of the factors binding to HoxC-WT the probe was incubated with a high molar excess (100-fold) of different competitors. No binding was detected when HoxC-WT was competed with itself, indicating that the observed shifts are due to specific protein-DNA interactions (Figure 4.6A, compare lanes 1 and 2). A similar result was obtained when the HoxP-WT probe was used as a competitor (Figure 4.6A, lane 6). A mutant form of the *Hoxb-4* promoter site failed to compete, however (HoxP-M, Figure 4.6A lane 7). The related HoxTF site from the *myogenin* gene binds only HoxTF and not YY1 (Myo-WT, Figure 4.5A, lane 3) and predictably, Myo-WT competed for the HoxTF band only (Figure 4.6A, lane 8) but a mutant form did not (Myo-M; Figure 4.5A, lane 4 and Figure 4.6A, lane 9). To verify the identity of the putative YY1 band HoxC-WT was competed by a well characterised YY1 consensus binding-site (Oligo 5, Ellis *et al.*, 1993; Dr. J. Ellis, personal communication) and supershifted by a specific antibody (Dr. R. J. Schwartz, Dept. of Cell Biology, Baylor College of Medicine, Houston, Texas. Data not shown).

These experiments demonstrate that a HoxTF/YY1 binding-site is present within CB1 of the *Hoxb-4* intron and is analogous to that previously described for the *Hoxb-4* promoter. The binding-sites for HoxTF and YY1 overlap. However, both factors may be competed individually indicating that are capable of binding in an autonomous manner. Comparison of the binding-sites from *Hoxb-4, myogenin* and *MyoD1* has allowed us to establish a consensus binding-site for HoxTF that is GCCA[G/T]T[G/C][G/C]C and is shown in Figure 4.5B (see also Gutman *et al.*, 1994). By comparing the relative strengths of shifts derived from the various probes it is possible to arrange the binding-sites in increasing order of HoxTF-binding affinity; *Hoxb-4* intron < *Hoxb-4* promoter < *myogenin* < *MyoD1*.

4.4 The HoxTF/YY1 binding-site is required for efficient transgene expression

The previously described deletion construct Δ514-606 (Figure 4.2) eliminated the CB1 sequences from region C. This deletion had a profound effect on the expression pattern, compared to that derived from the wild-type region C transgene. I have shown by *in vitro* analyses that the only detectable binding-site within this area is for HoxTF/YY1. Therefore, I wished to test whether the HoxTF/YY1 binding-site had a *cis*-regulatory role in transgenic mice and if its specific removal could reproduce the deletion phenotype of the Δ514-606 construct.

To investigate this, a construct was generated that carried a mutation in the HoxTF/YY1
The cloning steps used to make this construct (MH/Y) are described in the Materials and Methods Chapter (2.3.1). The 4bp mutation introduced into the HoxTF/YY1 site (GCCATTGCC to GCACGGGCC) was identical to that seen in Figure 4.5 (HoxC-M2, lane 7). When the wild-type probe (HoxC-WT) was incubated with an excess of cold mutant oligonucleotide (HoxC-MH/Y) in EMSA experiments, no competition of the HoxTF or YY1 binding activities was observed (compare lane 5 with lanes 1 and 2 in Figure 4.6A). This confirms that at the level of sensitivity of the in vitro assay, MH/Y represents a null mutation for HoxTF and YY1 binding.

The MH/Y reporter construct was used to produce transgenic mice which were analysed as transient F₀ embryos between 10.25 and 12.5 dpc. Of the sixteen transgenic embryos generated with this construct, ten expressed the reporter in a consistent manner and the results of this analysis are presented in Figure 4.7. The observed expression pattern was similar to that obtained with the deletion mutant Δ514-606 (Figure 4.2) with no staining present in the mesoderm. Expression was restricted to the nervous system but with a more extensive distribution than that of the Δ514-606 construct. These results show that the HoxTF/YY1 binding-site within CB1 is absolutely required for region C-directed mesodermal expression and for efficient expression of the transgene in the nervous system. Assuming that the MH/Y mutation completely abolishes the binding of HoxTF and YY1, as is suggested by the in vitro studies (Figures 4.5 and 4.6), these data also indicate the presence of a further regulatory element(s), in addition to the HoxTF/YY1 binding-site, that was deleted or perturbed in the Δ514-606 construct. Another possibility is that the MH/Y mutation creates a cryptic binding-site for another regulatory factor, the activity of which leads to additional expression within the nervous system in comparison to Δ514-606. These latter two points are discussed at greater length in the Summary and Discussion section (4.9) at the end of this Chapter.

At 10.25 dpc activity of the MH/Y construct was initially observed in the neural tube, running from the posterior tip to an anterior limit slightly rostral to the forelimb bud (Figure 4.7b). This boundary of neural tube expression is more posterior than that seen with the WT region C construct and the levels are much weaker. Therefore, not only is the HoxTF/YY1 binding-site required for appropriate levels of expression in the nervous system, but also for attaining the proper anterior boundary of region C-directed expression at the level of the spinal cord/hindbrain boundary. By 11.5 dpc an underlying pattern of ventral staining was apparent within the neural tube with a more widespread pattern at axial levels between the limb buds (Figure 4.7c-d). Staining was particularly evident at the level of the hindlimb and was visible within the neural tube, dorsal-root ganglia and ventral motor nerves innervating the developing limb buds. However, the pattern of neural staining is patchy and represents only a subset of that specified by region C, as described in Chapter 3. At 12.5 dpc the pattern observed was similar to that seen at earlier stages but more diffuse (Figure 4.7e-f). At this
Figure 4.7: *In vivo* requirement for the HoxTF/YY1 binding-site in CB1

The HoxTF/YY1 site in CB1 of region C is necessary for establishment of mesodermal and correct neural expression. The β-galactosidase expression pattern observed in transient transgenic embryos carrying the construct MH/Y is illustrated. A schematic of the MH/Y construct and of the parental region C construct (WT) is shown below. The legend to the reporter is as described in Figure 3.3. MH/Y carries a 4bp mutation in the HoxTF/YY1 site located within CB1 of region C and does not bind HoxTF or YY1 (Figure 4.6). Exp. denotes the total number of positively stained embryos obtained that showed a consistent pattern of expression. Tg. denotes the total number of transgenic embryos.

(a) Shows a lateral view of a 11.0 dpc transgenic embryo carrying the WT region C construct for comparison.

(b) Lateral view of a 10.25 dpc transgenic embryo carrying construct MH/Y. Staining can be seen in the neural tube up to a level slightly anterior to the forelimb bud. This is more posterior than that of the WT region C construct.

(c) Lateral view of an 11.5 dpc transgenic embryo carrying construct MH/Y. An underlying pattern of ventral staining can be seen within the CNS (v). A more widespread pattern of neural tube staining can be seen at levels between the limb buds.  

(d) Shows a transverse section through the neural tube of the same embryo, taken at the level of the hindlimb bud (lb). Patchy expression can be seen throughout the neural tube, including the floor plate (fp). Staining is also evident within the dorsal root ganglia (d) and ventral motor neuron tracts (n) of the peripheral nervous system.

(e) and (f) Show lateral and dorsal views of a 12.5 dpc transgenic embryo carrying construct MH/Y. Expression is predominantly restricted to the central nervous system with an anterior limit slightly rostral to the forelimb bud. Note the biphasic distribution of β-galactosidase staining in the neural tube, indicated by the hollow arrows.
stage, CNS staining adopted a biphasic pattern with strong patches in the spinal cord at axial levels adjacent to the fore- and hind-limb buds. These data illustrate that the HoxTF/YY1 binding-site is largely required for the activity of region C throughout its neural expression domain.

The transgenic results from the MH/Y construct clearly demonstrate that the HoxTF/YY1 binding-site is a critical \textit{in vivo} \textit{cis}-regulatory component of region C. Although the patterns of expression observed with the MH/Y construct are not identical to those seen with the \( \Delta 514-606 \) deletion mutant, the effects of mutating this element in the context of the intact enhancer are widespread and indicate that it is essential for many aspects of region C-directed expression. These include activity within the mesoderm, and the levels and tissue distribution of expression within the central and peripheral nervous systems. In the context of the transcriptional regulation of the \textit{Hoxb-4} gene the role of region C activity within the nervous system is unclear, since in this respect it seems to functionally overlap with the 3' \textit{Hoxb-4} neural enhancer region A (Whiting et al., 1991). However, as all aspects of \textit{Hoxb-4} mesodermal expression require the activity of region C it may be logically concluded that the HoxTF/YY1 binding-site is an essential component of this pattern. The results of my studies thus far have demonstrated that evolutionary sequence comparison, in conjunction with transgenic reporter gene analysis, is an useful approach for the identification of noncoding regions that function in \textit{cis}-regulation. Once such regions have been identified, the use of \textit{in vitro} EMSA analyses have proved valuable for the rapid identification of functional binding-sites within them.

\section*{4.5 HoxTF is the major \textit{trans}-activating component \textit{in vivo}}

Our previous work on the \textit{Hoxb-4} promoter HoxTF/YY1 binding-site had enabled us to define mutations that specifically abolish the binding of either factor. These data have shown that HoxTF and YY1 bind independently to overlapping binding-sites and do not require a cooperative interaction. Furthermore, we have shown that only HoxTF, and not YY1, is able to mediate \textit{trans}-activation of a reporter gene driven by the \textit{Hoxb-4} promoter in cultured Neuro2a cells (Gutman et al., 1994). To investigate the individual roles of HoxTF and YY1 binding \textit{in vivo}, I engineered constructs that carried specific mutations in the basic WT reporter and assayed their effect in transgenic mice. The cloning steps used to generate these constructs (MYY1, MHTF1 and MHTF2) are described in the Materials and Methods Chapter (2.3.1).

Oligonucleotides carrying the various mutations were tested as competitors in EMSA experiments to verify their binding characteristics. The wild-type probe (HoxC-WT) was
labelled with [γ32P]-ATP and EMSA experiments were carried out as described previously. Whole-cell extract from the mouse neuroblastoma cell-line Neuro2a, containing both the HoxTF and YY1 binding activities, was used as a source of protein. HoxC-WT was incubated in the presence of a large molar excess (100-fold) of the various unlabelled, mutant oligonucleotides. The results of these analyses are shown in Figure 4.6. The YY1 mutant (HoxC-MYY1) competed out the HoxTF but not the YY1 binding activity (compare lane 3 with lanes 1-2 in Figure 4.6A), indicating that it bound only HoxTF and not YY1. Conversely, with a HoxTF mutant (HoxC-MHTF2; lane 4, Figure 4.6A) the YY1 but not the HoxTF band was competed. A second, incomplete HoxTF mutant was also analysed (MHTF1, binding data not shown; Gutman et al., 1994) which retained approximately 15-20% of HoxTF binding activity compared to that of HoxC-WT.

The expression patterns derived from transgenic mice carrying the MYY1, MHTF1 and MHTF2 mutant constructs are illustrated in Figure 4.8. The YY1 mutant reporter (MYY1) displayed a similar pattern to that of WT but there were significant differences in the levels of mesodermal expression (Figure 4.8a-b). All seven of the positively stained embryos obtained with MYY1 showed a consistent pattern of β-galactosidase activity, although there was clear ectopic, neural expression in the head region of four of them. Expression within the neural tube was normal but the levels of staining within the somitic and lateral mesoderm were much weaker than those of WT. Staining was visible within the somitic mesoderm, albeit weakly, up to the level of the forelimb, becoming stronger more posteriorly. These results indicate that the binding of YY1 to the HoxTF/YY1 site may have a specific role in the transcriptional activation of region C within mesodermal tissues. The presence of anterior neural expression in some of the embryos (4/7) may point to an additional, normally repressive, role of YY1 in this region. Another possibility which can not be excluded is that the introduced MYY1 mutation also has a subtle effect on the binding affinity of HoxTF. This could explain the ectopic anterior expression seen in terms of position and copy number effects at the site of transgene integration, more evident when HoxTF binding is perturbed, whilst a reduction in HoxTF binding could cause a corresponding reduction in the levels of reporter gene activity, most evident in the mesoderm. The EMSA analysis of the YY1 mutation supports this hypothesis to some extent, since the binding of HoxTF to the wild-type sequence (HoxC-WT) is competed less efficiently by the HoxC-MYY1 than by HoxC-WT itself (compare lanes 1 to 3 in Figure 4.6). The effect of the MYY1 mutation on HoxTF binding has not been quantified, however, it is probably in the order of a 5-10% reduction compared to wild-type levels.

Two different mutations were introduced into the HoxTF binding-site, MHTF1 and MHTF2. MHTF1 is more conservative and less drastic than MHTF2 and retains approximately 20% of HoxTF binding activity compared to that of wild-type. MHTF2 does not appear to bind
HoxTF binding activity is essential for establishment of mesodermal and correct neural expression by region C. The β-galactosidase expression patterns observed in transient transgenic embryos carrying constructs MYY1, MHTF1 and MHTF2 are illustrated. A schematic of the reporter constructs and of the parental region C construct (WT) is shown below. The legend to the reporters is as described in Figure 3.3. The relative HoxTF or YY1 binding capacities of each reporter is shown above it (+ = binds, - = no binding, +/- = partial binding). Exp. denotes the total number of positively stained embryos obtained that showed a consistent pattern of expression. Tg. denotes the total number of transgenic embryos.

(a) and (b) Show lateral and dorsal views of an 11.5 dpc transgenic embryo carrying the MYY1 construct that binds HoxTF but not YY1. The expression pattern obtained is similar to that of the WT region C construct. Staining within the nervous system is normal but that within the lateral mesoderm (Im) and somitic mesodermal (sm) is much weaker.

(c) and (d) Show lateral and dorsal views of a 12.5 dpc transgenic embryo carrying the MHTF1 construct that binds YY1 and also HoxTF at about 15-20% of its wild-type level. Mesodermal staining is absent. β-galactosidase activity is present within the neural tube and more laterally within the dorsal root ganglia of the peripheral nervous system (d).

(e) and (f) Show lateral and dorsal views of a 12.5 dpc transgenic embryo carrying the MHTF2 construct that only binds YY1 and not HoxTF. Patchy expression is limited to the neural tube and levels of β-galactosidase activity are comparable to those seen with the MH/Y construct (Figure 4.7).
HoxTF at all, at least under the conditions of the EMSA experiments. Both MHTF1 and MHTF2 reporters showed some activity in the neural tube, although at different levels, but neither was consistently expressed in the somitic or lateral mesoderm. MHTF2 activity in the neural tube was low in all three of the positively stained transgenic embryos obtained and none exhibited mesodermal activity. The level of β-galactosidase activity exhibited by MHTF2 was comparable to that of the double HoxTF/YY1 mutation initially tested (MH/Y, Figure 4.7). In contrast, expression levels of MHTF1 were generally higher and more variable than those of MHTF2 and staining was detected throughout the neural tube and within the dorsal root ganglia of the PNS. In addition, five out of the sixteen positively stained embryos exhibited some mesodermal expression at a very low level. This variability in expression levels obtained with MHTF1 was probably due to transgene position and copy number effects, more evident in a partial mutant.

The results of these selective mutational experiments indicate that activation of region C via the HoxTF/YY1 binding-site in CB1 is largely mediated through the HoxTF binding activity. YY1 may play some role in transcriptional activation within the mesoderm, although the effects of the introduced YY1 mutation may be explainable in terms of a reduction in HoxTF binding affinity. A more specific role for YY1 in region C-directed transcriptional regulation, such as in discrete tissue-specific or temporal aspects of expression, cannot be discounted and this is considered at greater length in the Summary and Discussion (4.9).

4.6 The HoxTF binding-site activates transcription in CAT assays

The results from my transgenic experiments have clearly shown that the HoxTF binding-site is required for normal region C activity in vivo. The results concerning YY1 were ambiguous, however. It was not apparent whether YY1 acts as an transcriptional activator or repressor as has been previously documented (Lee et al., 1991, 1994; Shi et al., 1991; Riggs et al., 1993), or whether it plays no significant role in transcriptional regulation by region C. To further investigate the effects of HoxTF and YY1 binding through the Hoxb-4 intron HoxTF/YY1 site, we designed a series of chloramphenicol acetyltransferase (CAT) reporter constructs containing wild-type and mutant binding-sites and assayed their activities in cultured cells. Similar experiments using the HoxTF/YY1 site from the Hoxb-4 promoter have shown that HoxTF is able to trans-activate the Hoxb-4 minimal promoter, whilst YY1 seems to have no significant effect (Gutman et al., 1994).

The basic reporter construct used consisted of Hoxb-4 minimal promoter sequences (-32 to +31) driving the bacterial CAT gene, upstream of which various oligonucleotide binding-sites were cloned. The wild-type construct (WT) contained one or three head to tail copies of
the HoxTF/YY1 binding-site (bp 510-557 of region C). MH/Y, MHTF2 and MYY1 contained three head to tail copies of the same oligonucleotide but carried the various mutations used in the transgenic analyses and described previously (Figure 4.6). These experiments were performed by Dr Alejandro Gutman in Dr. P. W. J. Rigby's Laboratory as described in the Materials and Methods Chapter (2.5) and in Gilthorpe et al. (manuscript in preparation). The results are summarised in Figure 4.9.

The wild-type construct carrying three copies of the HoxTF/YY1 binding-element (WT 3X), with both HoxTF and YY1 sites intact, increased the CAT activity of the minimal promoter (MP) by about five-fold. In contrast, the same wild-type element when presented as a single copy had only a slight effect (WT 1X). A mutation that abolishes the binding of both HoxTF and YY1 reduced transcription to the basal level of MP (MH/Y). The same effect was achieved by mutating the HoxTF site alone (MHTF2). Finally, a mutation specific for YY1 had no effect (MYY1). These results indicate that HoxTF acts at the level of transcription and is able to activate a minimal promoter when presented with multiple binding-sites of the Hoxb-4 intron type. YY1 seems unable to do so, at least in the context of this assay, and selectively mutating the YY1 binding-site had no effect on the level of activation. The relatively weak level of activation seen with a single copy of the binding-site (WT 1X) probably reflects the relatively low-affinity of HoxTF for the site in the Hoxb-4 intron. Gutman et al. (1994) have reported a much higher level of activation by a single copy of the Hoxb-4 promoter site (6- to 7-fold over the promoter alone). This probably reflects the higher affinity of HoxTF for the promoter site (compare lanes 1 and 2 in Figure 4.5).

4.7 Activation by HoxTF in vivo

I have demonstrated that HoxTF is necessary for trans-activation of a region C-lacZ reporter construct in transgenic mice. Multimerised HoxTF binding-elements are also able to activate the Hoxb-4 minimal promoter in cultured cells. I next wished to investigate whether the same element was able to drive expression of a reporter gene in transgenic mice. Two transgenic reporter constructs were made that carried one or three copies of the HoxTF binding-element (MYY1, Chapter 4.6 and Figure 4.9) upstream of hsp68-lacZ. The cloning steps used to generate these constructs (1XHTF and 3XHTF) are described in the Materials and Methods Chapter (2.3.1).

1XHTF and 3XHTF were used to produce transgenic mice which were analysed as transient F0 embryos at 11.5 to 12.0 dpc. The results of these analyses are presented in Figure 4.10. Of the nine transgenic embryos generated with the 1XHTF construct, six expressed the transgene in a consistent manner. β-galactosidase activity was most evident within elements
Figure 4.9: The HoxTF binding-site activates transcription in CAT assays

A summary of cell transfection experiments used to investigate the transcriptional effects of HoxTF and YY1 binding on the *Hoxb-4* minimal promoter. A schematic of the recombinant reporter constructs tested is shown. Each construct carries one (1X) or three (3X) head to tail copies of various wild-type or mutant binding-sites, upstream of the *Hoxb-4* minimal promoter sequences (-32 to +21) linked to CAT (MP). The oval represents the CAT (chloramphenicol acetyltransferase) reporter plasmid pBLCAT3 (Luckow & Schutz, 1989). The broken arrow above the *Hoxb-4* promoter (large box) indicates the location of the transcription start site.

WT 1X contains one copy of an oligonucleotide that binds HoxTF and YY1 encompassing sequences from bp 510-557 of region C; WT 3X contains three head to tail copies of the same oligonucleotide; MHTF2, MYY1 and MH/Y carry mutations in either the HoxTF or in the YY1 sites or in both. The sequence of the mutations is as shown in Figure 4.6. The small boxes are a schematic representation of the binding-sites: HoxTF (left) and YY1 (right); white boxes indicate wild-type sequences and shaded boxes, mutated ones. The arrows indicate the orientation of the oligonucleotides.

Reporter recombinants (16 μg) were were co-transfected with 4 μg of the internal control plasmid RSV-βgal into the mouse neuroblastoma cell-line Neuro2a, as described in the Materials and Methods Chapter (2.5). Cell extracts were prepared and CAT activities were measured, corrected by subtraction of background levels of the promoterless vector pBLCAT3 and normalised with respect to MP (taken as 1). CAT activity values are the average from three independent transfections and are shown with their corresponding standard deviations.
### Reporters

| Reporter | 
|---|---|
| MP | Hoxb-4 |
| WT 1X | Hoxb-4 |
| MH/Y | Hoxb-4 |
| MHTF2 | Hoxb-4 |
| MYY1 | Hoxb-4 |

### Relative CAT Activities

- **MP**
  - CAT: 1.00

- **WT 1X**
  - CAT: 1.55 ± 0.25

- **MH/Y**
  - CAT: 5.23 ± 0.61

- **MHTF2**
  - CAT: 0.83 ± 0.04

- **MYY1**
  - CAT: 5.10 ± 0.49
Figure 4.10: The HoxTF binding element is sufficient to drive a subset of the *Hoxb-4* neural expression pattern

The patterns of β-galactosidase activity derived from HoxTF binding-site reporter constructs are shown. 1XHTF and 3XHTF carry one or three head to tail copies of the HoxTF binding element, respectively. The oligonucleotide containing this element covers bp 510-557 of region C and is identical to that previously described (MYY1, Chapter 4.6 and Figure 4.9). A schematic of both constructs is shown below. The arrows indicate the orientation of the oligonucleotides. The small black and white boxes above the arrows represent the binding-site and follow the legend in Figure 4.9. Exp. denotes the total number of positively stained embryos obtained that showed a consistent pattern of expression. Tg. denotes the total number of transgenic embryos.

(a) Shows a lateral view of a 12.0 dpc transgenic embryo carrying the 1XHTF construct. Staining due to β-galactosidase activity can be seen in the peripheral nervous system (p). Additional ectopic expression can be seen in the head.

(b) Lateral and (c) dorsal views of an 11.5 dpc transgenic embryo carrying the 3XHTF construct. The pattern of β-galactosidase staining obtained is stronger but essentially the same as that of 1XHTF. Strong CNS staining can be seen in the spinal cord, although weaker staining extends more rostrally in to the midbrain. Staining is also visible within the proximal region of the forelimb (li) and in the eye of several embryos. Transverse sections, lightly counterstained with eosin, taken through the same embryo are shown in (d-f). (d) and (e) show dark-field images (10X and 40X magnification, respectively). The plane of section is indicated by the arrowheads (pi) in (b). Under dark-field conditions the blue stained regions fluoresce pink. Staining can be seen throughout the neural tube but concentrated within the ventral motor horns. Staining in the dorsal root ganglia (dg) is generally restricted to the medial edge. Ventral motor neuron tracts can be seen innervating the proximal region of the forelimb (f) where there is a strong patch of expression (li). (f) Shows a high power (100X magnification) bright-field image of the ventral neural tube. The picture covers the white boxed region in (e). Note the clear axonal morphology of the positively stained cells, such as the one circled. (fp) floor plate, (dg) dorsal root ganglion containing a positively stained cell (boxed).
1XHTF

3XHTF

Exp. | Tg.
---|---
6   | 9   
7   | 10  

- HoxTF
- YY-1

hep 68

lacZ

SV40 pA

p2

li

f

dg

fp

pl
of the peripheral nervous system, notably the dorsal root ganglia and ventral motor neuron tracts (Figure 4.10a). Some expression was seen in the neural tube although this was relatively weak (data not shown). The consistent pattern of neural staining was also accompanied by varying degrees of ectopic expression within the brain and in the mesenchyme of the head and trunk. Three copies of the HoxTF binding-element (3XHTF) generated a pattern of expression very similar to that seen with the 1XHTF construct. However, reporter gene activity was generally higher and less variable ectopic expression was evident, with no obvious mesenchymal staining (Figure 4.10b-c). Seven 3XHTF embryos were obtained that showed a consistent pattern of expression, from a total of ten transgenics. Staining within the CNS was evident along the length of the spinal cord and more diffusely throughout more anterior neural structures, such as the hindbrain and midbrain, in which there was no obvious anterior limit of expression. PNS staining was visible in all of the spinal ganglia and the associated motor nerves and within some of the cranial ganglia, although this latter aspect of the pattern seemed to vary from embryo to embryo. Several other sites of expression were seen in some, but not all, of the positively stained embryos including; the developing neural layer of the eye (3/7) and the proximal region of the limb (4/7). Transverse sections of these embryos show that spinal cord staining is predominantly restricted to the ventral region of the neural tube (Figure 4.10d-f). Positively stained cells have an axonal morphology and are mainly located within the ventral motor horns but are also present in the floor plate. The ventral motor neuron tracts of the PNS are strongly stained, as are occasional cells within the dorsal root ganglia.

These data show that an isolated HoxTF binding-site is capable of behaving as a positive regulatory element in vivo and is sufficient to drive a limited pattern of expression within the central and peripheral nervous systems during mid-embryogenesis. When multimerised, this same element produces a more refined pattern of neurally-restricted expression. The HoxTF binding-site seems to be essential for the activity of Hoxb-4 in the mesoderm but it is not not sufficient for mesodermal expression and, therefore, HoxTF must interact with other positively-acting factors to achieve this. Contrastingly, whilst HoxTF-element-driven expression may be considered to be a subset of that dictated by region C, it also extends more rostrally to the normal region C boundary at the junction of the spinal cord/hindbrain. Thus, negatively-acting elements must also be present within region C that can restrict this activity.

4.8 HoxTF is widely distributed and binds to DNA as a heterodimer

I have established that an isolated HoxTF binding-element is capable of driving spatially-restricted reporter gene expression in transgenic mice. Furthermore, this element is required for two aspects of Hoxb-4 regulation, the activity of the promoter and the activity of the
downstream enhancer, region C. The HoxTF target sequences we have identified do not appear to correlate with those for any previously identified transcription factors (Gutman et al., 1994). Therefore, it was important to begin further characterisation of HoxTF with a view to cloning it.

Figure 4.11A shows the results of EMSA experiments using the HoxC-WT oligonucleotide as a probe to look at the spatial distribution of the HoxTF binding activity in a variety of cell-line and embryonic tissue extracts. The results of these analyses show that HoxTF has a wide distribution throughout the embryo and in the cell-line extracts tested. In the 10.5 dpc embryo HoxTF is present both within structures that express the endogenous Hoxb-4 gene, such as the hindbrain and branchial arches (Hoxb-4 expressing cells contribute to the formation of the fourth branchial arch), but also within those that do not, such as the limb buds, liver, heart and forebrain. Additionally, the HoxTF binding activity is detectable in both 8.5 dpc and 14.5 dpc embryo extracts (data not shown). HoxTF is also found in a variety of mouse (Neuro2a; C2 myoblasts and myotubes; F9 EC), and the human HeLa, cell-lines. It was not detected in the T-lymphocyte derived JM cell-line, although no control was available for the integrity of this extract. The levels of HoxTF generally bear the same relationship to those of YY1 though there are some potentially interesting exceptions. For example, the levels of HoxTF are much higher in F9 EC cells when compared to those of HeLa cells (compare lanes 10 and 11 in Figure 4.11A). Another notable feature is that there are small but noticeable differences in the mobility of the HoxTF/DNA complex with different extracts. For example, a slower migrating complex is present in Neuro2a cells (lane 7) than in F9 EC cells (lane 11). This may well indicate that HoxTF exists in a number of different forms in different tissues.

Figure 4.11B shows the results of a UV-crosslinking analysis designed to investigate the number of polypeptide species associated with the HoxTF binding activity. This experiment was performed by Dr Alejandro Gutman in Dr. Rigby's Laboratory as described in the Materials and Methods Chapter (2.5). The crosslinking technique relies on the covalent interaction of proteins bound to DNA via the base analogue 5-bromo-deoxyuridine (BrdU), when exposed to ultraviolet light (Chodosh et al., 1986). A HoxC-WT probe (bp 524-550 of region C, Figure 4.5) was synthesised that contained the substituted base BrdU, instead of deoxythymidine (T). The labelled probe was used in preparative EMSA reactions, using Neuro2a extract as a source of protein, in the presence of the specific (HoxC-WT), or a mutant (MHTF2), competitor. Samples were irradiated with UV light to crosslink bound proteins and the products of the reactions were resolved by SDS-PAGE. Dr. Gutman detected two major bands of approximately 115KD and 90KD and one minor band of approximately 85KD. These three polypeptide species are specific as they are competed out by an excess of cold wild-type oligonucleotide (HoxC-WT) but not by a HoxTF mutant.
Figure 4.11: HoxTF is widely distributed and binds to DNA as a heterodimer

(A) Shows the results of EMSA experiments used to determine the distribution of the HoxTF binding activity in a variety of embryonic tissues and cell-lines. The HoxC-WT probe was labelled with $[^{32}P]ATP$ as described in the Materials and Methods Chapter (2.4). Whole-cell extracts ($15-20 \mu g$) were used as a sources of proteins and are indicated above each lane. Extracts were prepared from the dissected tissues of 10.5 dpc mouse embryos or cultured cell-lines as described in Chapter 2.4.2; B. arches = branchial arches, C2 Mb = undifferentiated C2-myoblasts, C2 Mt = differentiated C2-myotubes. Other extract names are self explanatory. The products of the binding reactions were resolved on a 5% (w/v) polyacrylamide gel. The two major retarded bands corresponding to either HoxTF or YY1 interactions are indicated. FREE indicates the unbound probe.

(B) Shows the results of a UV-crosslinking experiment used to determine the number and apparent size of polypeptide species binding to the HoxTF site. The end-labelled HoxC-WT probe used is identical to that described in (A), except that the modified base 5-bromo-deoxyuridine substitutes deoxythymidine (T). Whole-cell extract from the mouse neuroblastoma cell-line Neuro2a was used as a source of protein. A preparative binding reaction was made as described in Material and Methods Chapter (2.4.4), in the presence of the indicated competitors. The HoxC-MHTF2 oligonucleotide is described in Figure 4.6 and binds YY1 but not HoxTF. Samples were irradiated with UV light for 30 minutes and resolved on an 8% (w/v) SDS-polyacrilamide gel. The relative positions of molecular weight standards are shown, together with their sizes in kilo daltons (KD). The arrows point to the detected bands.
(HoxC-MHTF2, Figure 4.6). The estimated apparent molecular weights of the three species are probably slightly high due to the presence of the HoxC-WT probe (approximate Mr 18kD) which may influence the migration of the polypeptides. The most likely interpretation of this result is that HoxTF binds to DNA as a heterodimer and that the two major bands correspond to the two different monomers. The minor band might represent a modified form of one of the monomers or a partial degradation product.

4.9 Summary and Discussion

The comparative sequence analysis described in Chapter 4 has revealed the presence of a conserved block of homology (CB1) present in all PG-4 Hox gene introns examined. The highest degree of sequence identity to the mouse Hoxb-4 intron was seen in the other homologous sequences, i.e chicken and Fugu Hoxb-4. Deletion of CB1 from the mouse intron in construct Δ514-606 showed that it contains sequences necessary for all aspects of region C-directed expression, namely in the mesoderm and central and peripheral nervous systems. Some staining due to residual β-galactosidase activity was detected within the ventral neural tube of 50% (6/12) of the transgenic embryos generated with this construct. It is not known whether this pattern of residual activity seen with the Δ514-606 construct derives specifically from elements within region C, or from an artifactual response of the hsp68 promoter-lacZ reporter construct. The hsp68 promoter fragment has been shown to have no detectable basal activity in transgenic mouse embryos (Kothary et al., 1989) but it can be activated in distinct patterns by defined heterologous enhancer elements (e.g. Tuggle et al., 1990; Whiting et al., 1991; Becker et al., 1996). Several researchers, however, have described a pattern of consistent but ectopic expression in the ventral neural tube with hsp68 promoter constructs. This has been attributed to an element in the hsp68 promoter that is able to interact with enhancer elements juxtaposed to it (Gérard et al., 1993; Logan et al., 1993, Song et al., 1996). At some frequency we have seen this pattern of expression with most of the constructs described. It is therefore likely that the ventral neural tube staining we have observed is not an effect of residual region C-directed enhancer activity.

Over the 96bp length of CB1 the most highly conserved sequences lie towards the 3' end, corresponding to bp 571-601 of region C. All sequences examined share a strong identity to the mouse Hoxb-4 intron over this 30bp region (70-90%) and possess two common A/T-rich motifs. These motifs correspond to potential binding-sites for known transcriptional regulators of Hox/HOM gene expression, namely homeodomain (Motif III), and hunchback and caudal-related (Motif IV) proteins (Zuo et al., 1991; Margalit et al., 1993). These data were potentially significant as they implied that the sequences of CB1 may have a conserved regulatory function, maintained throughout the evolution of the Hox complexes. In
Drosophila, hunchback activity is required for the earliest phases of Deformed activation at the cellular blastoderm stage (Jack et al., 1988; Jack & McGinnis, 1990). In the mouse, the caudal-related gene Cdx1 is required for the appropriate expression of Hox genes within the somitic mesoderm. Disruption of Cdx1 leads to anterior homeotic transformations of the axial skeleton, concomitant with a posterior shift in the expression domains of Hox genes in the somites by a distance of one or two segmental units (Subramanian et al., 1995). Although the expression domain of Hoxb-4 has not been documented in Cdx1-/- mice, it seems highly plausible that its mesodermal expression domain will be similarly affected in light of the wide range of homeotic transformations described. The potential regulatory role of Cdx1 in Hox gene expression is of particular interest when considering the capacity of region C in defining the somitic expression domain of Hoxb-4. Deletion of the most conserved sequences of CB1 in construct A558-598, including the potential Cdx binding-motif, had no obvious effect on the overall expression pattern of the transgene. Given the extremely weak anterior boundaries of expression produced by the region Clhsp68 promoter construct, it is possible that there is a shift in the boundary of somitic expression which is not observable in my assay. However, EMSA analyses of these sequences have shown no specific binding of factors with extracts from 8.5-14.5 dpc embryos and from a variety of cell-lines (data not shown). In light of this evidence I cannot suggest a definitive regulatory function for the most conserved sequences of CB1, although their strong evolutionary conservation remains intriguing. It is possible that CB1 is required for aspects of Hoxb-4 expression that are not been detectable with my transgenic assay, such as levels of expression, or the appropriate temporal activation of Hoxb-4 in its natural genomic context as part of the Hoxb cluster.

The transgenic and EMSA analyses of CB1 have revealed the presence of overlapping binding-sites for HoxTF and YY1. We have previously identified a similar binding-site in the 5'-untranslated region of the Hoxb-4 promoter (Gutman et al., 1994). In this context the HoxTF site can mediate transcriptional activation of the Hoxb-4 minimal promoter. The intron site behaved like that of the promoter in EMSA experiments and competed for the same binding activities. A transgene carrying a mutation in the HoxTF/YY1 site (MH/Y), that abolished the binding of both factors, exhibited a similar pattern of activity to the CB1 deletion construct (A514-606). This indicates that, in addition to the promoter, the HoxTF/YY1 site is a critical regulatory component of the region C enhancer. Interestingly, the HoxTF/YY1 site is poorly conserved amongst the PG-4 intron sequences examined. The greatest degree of conservation is seen with the chicken Hoxb-4 sequence, which matches the HoxTF consensus at 7/9 bases and contains a conserved core (ACCATTCGT). The equivalent sequences in the Fugu intron match the defined consensus for HoxTF at 7/9 bases (GCCTGTGGG), but do not compete for the mouse binding activity in EMSA experiments (data not shown). One poignant question that arises from this information relates to a possible conserved role of HoxTF in the regulation of Hoxb-4 expression. Is HoxTF a
fundamental regulator of \textit{Hoxb-4} throughout the vertebrate lineage? Chicken \textit{Hoxb-4} contains a region C-like intronic enhancer, able to mediate expression in some of the same posterior mesodermal and neural tissues in transgenic mice, though unable to specify the appropriate anterior boundary of somitic expression (Morrison \textit{et al.}, 1995). In an analogous manner to my experiments, CB1 is also essential for mesodermal and posterior neural tube activity of a chicken \textit{Hoxb-4} transgene. Conversely, the intron of \textit{Fugu Hoxb-4} is incapable of specifying a region C-like pattern (Morrison \textit{et al.}, 1995) and does not contain a canonical HoxTF/YY1 site. It is possible that the \textit{Fugu Hoxb-4} intron contains a functional HoxTF site but that the mouse factor is sufficiently divergent and unable to recognise it. However, the HoxTF/YY1 site in the \textit{Hoxb-4} promoter is absolutely conserved between mouse, chicken (Gutman \textit{et al.}, 1994) and \textit{Fugu} (data not shown), arguing against this. I think, therefore, that HoxTF represents a component of an evolutionarily conserved mechanism for \textit{Hox} gene regulation, but in the \textit{Fugu} intron the HoxTF site has moved or its role is functionally compensated for by other elements that are not recognised by factors in the mouse.

Whilst the CB1 deletion mutant $\Delta 514-606$ exhibited very limited expression in the ventral neural tube, the MH/Y mutation gave a more extensive pattern of neural staining. Several reasons maybe put forward to explain this difference. Firstly, additional transcription factor binding-sites may be present in CB1 that were not detected in our EMSA experiments. These sites would have been removed in $\Delta 514-606$ but not in MH/Y. A second possibility is that the $\Delta 514-606$ deletion altered the relative spacing of flanking regulatory elements. This may have rendered them unable to interact with one another, producing the more severe phenotype. A third possibility is that the introduced MH/Y mutation generates a cryptic binding-site for another regulatory protein. In light of the transgenic data this third possibility seems to provide the most likely explanation, at least in part. The MH/Y transgene showed a pattern and progression of $\beta$-galactosidase activity similar to that previously described for a retinoic acid response element (RARE) driven transgene (Balkan \textit{et al.}, 1992; Colbert \textit{et al.}, 1995). This reporter consisted of three copies of a RARE half-site cloned upstream of the herpes simplex virus \textit{thymidine kinase} promoter and coupled to \textit{lacZ}. In the developing spinal cord expression of the transgene was detected throughout the AP axis at 10.5 \textit{dpc} and then became restricted to two distinct posterior and anterior domains at 12.5 \textit{dpc} (Colbert \textit{et al.}, 1995). Examination of the MH/Y mutation shows that it generates a sequence (GGGCCA) that matches the RARE half-site consensus ([G/A]G[G/T]TCA) at 5/6 positions. Retinoic acid receptors are known to activate transcription via target sequences composed of directly or palindromically repeated half sites (reviewed in Stunnenberg, 1993; Jansa & Forejt, 1996). No other obvious RARE half-sites are located in proximity to the one generated by the MH/Y mutation. Recent research has shown, however, that RARE half-sites can function when separated by large distances, i.e. up to 150bp apart (Kato \textit{et al.}, 1995). In light of these data it is plausible that the anomalous expression pattern seen with
MH/Y, compared to that of \( \Delta 514-606 \), may be partly attributed to a cryptic RARE. In support of this hypothesis, similar biphasic RARE-like patterns of expression were not seen with other constructs carrying different mutations in the HoxTF/YY1 site (MHTF1 and MHTF2). On the other hand, none of the HoxTF/YY1 mutation constructs demonstrated an identical pattern of expression to \( \Delta 514-606 \) which also implicates the involvement of other regulatory elements in the locality of the deletion.

Independent mutational analyses of the HoxTF/YY1 binding-site indicate that the major influence on transcriptional activation stems from HoxTF and not YY1 binding. A transgene carrying a selective mutation (GCCAGTGGC) that abolished only YY1 binding (MYY1), gave an equivalent pattern neural expression to the wild-type construct, though the levels of mesodermal activity were much reduced. In contrast, selective mutations that reduced or abolished HoxTF binding showed no mesodermal expression and varying degrees of activity within the nervous system. MHTF2 which gives no residual HoxTF binding in vitro, showed low levels of neural expression that were comparable to those of the complete HoxTF/YY1 mutant (MH/Y). Intermediate levels of HoxTF binding (15-20% of wild-type) drove much higher levels of neural activity as demonstrated by the incomplete HoxTF mutant construct MHTF1. Slightly reduced levels of HoxTF binding, as seen in the YY1 mutant construct (MYY1), showed only a reduction in the levels of mesodermal expression. These results suggest that the capability of HoxTF to direct tissue-specific expression via the region C enhancer is dependent on the relative strength of the HoxTF binding-site. Mesodermal expression may require much higher levels of HoxTF binding activity than does neural expression. These data may explain the reduced levels of mesodermal activity seen with the MYY1 transgene, due to a slight reduction in the affinity of HoxTF for the mutated site. The MYY1 mutation could affect the binding of HoxTF either directly, by creating a site for which HoxTF has a lower affinity, or by an indirect mechanism whereby YY1 binding increases the affinity of HoxTF for its target site. Such cooperativity between HoxTF and YY1 was not evident in any of the in vitro binding or CAT assay experiments. However, binding of YY1 to a site overlapping the serum response elements (SRE) within the c-fos promoter enhances the binding of serum response factor (SRF) to this site (Natesan & Gilman, 1995). This effect seems to stem from the ability of YY1 to bind to the major groove of the DNA and induce a conformational change, thereby enhancing the binding of SRF to the minor groove (Natesan & Gilman, 1993, 1995). Such a mechanism may also apply to the HoxTF/YY1 binding site, accounting for the reduction of HoxTF binding in the YY1 mutant.

Corroborating evidence for the role of HoxTF as a trans-activator of Hoxb-4 expression has been derived from cell-transfection experiments using isolated HoxTF/YY1 binding-elements. A single wild-type HoxTF/YY1 binding-element was able to marginally activate a
Hoxb-4 minimal promoter-CAT construct in the Neuro2a cell-line (1.5-fold above the basal level). When presented as three multimerised copies, the same element conferred approximately 5-fold activation. Similar experiments have shown that a single copy of the Hoxb-4 promoter site activates the same minimal promoter construct 7-fold, i.e. approximately five times more strongly than the intron site (Gutman et al., 1994). This correlates well with the relative affinities of HoxTF for the two sequences as judged by comparative EMSA experiments, the promoter site binding HoxTF approximately five times more effectively than the intron site, although the levels of YY1 binding are equivalent. Elements carrying mutations that abolished the binding of HoxTF showed no activation, whereas the YY1 mutant gave levels of expression equivalent to that of the wild-type site. These results confirm the transgenic data and show that HoxTF is capable of trans-activating the Hoxb-4 promoter. Regarding YY1, however, the results of the transfection experiments do not lend any clarification to a possible role in transcriptional regulation by region C. No significant change in the level of trans-activation was seen with the YY1 mutant construct as compared to wild-type. This does not agree with the transgenic experiments where the same mutation did have an effect on reporter gene expression in the mesoderm. Although weak, this effect highlights a shortfall of the cell culture system.

Somewhat surprisingly, a single copy of the intron HoxTF binding-element was able to activate the hsp68-lacZ reporter construct in vivo to define a limited pattern of reporter gene expression at 11.5-12.0 dpc. Reporter gene activity was consistently limited to regions of the neural tube and peripheral nervous system with the 1XHTF construct. With three copies of the same element (3XHTF), the expression pattern generated was stronger but retained the same tissue-specificity. Very little ectopic expression, attributable to positional effects at the site of transgene integration, was also a feature of multimerising the binding-element. This may indicate that HoxTF has a preferential tendency to auto-interact, rather than cooperate with other factors and could be important for region C interactions with the Hoxb-4 promoter. The highest levels of β-galactosidase activity detected with the HoxTF element construct were in the ventral neural tube and ventral motor neuron tracts, but expression extended rostrally from the spinal cord into the midbrain/forebrain region and positively stained cells were also visible in the proximal region of the limbs. Thus, elements within region C must act to repress these aspects of HoxTF-driven expression that are not normally associated with the activity of the enhancer. It is clear from our research that HoxTF is a necessary component for all aspects of region C-directed expression, particularly within the mesoderm. Despite the apparent ubiquity of the HoxTF binding activity, the HoxTF binding-element is only sufficient to confer expression within a subset of neural tissues. This may indicate a requirement for HoxTF to interact with other factors, or that the active transcriptional form of HoxTF requires some form of post-translational modification in order achieve mesodermal expression. It is interesting to note that the Hoxb-4 promoter alone is only able to drive a
limited pattern of ectopic expression within the midbrain (Whiting et al., 1991). Since the promoter contains a functional HoxTF site, negatively acting elements within the promoter region must be capable of suppressing the more posterior neural expression that I have described for the 1XHTF construct. Whether these negative elements represent those described by our cell culture analyses (Gutman et al., 1994) awaits further study.

Our UV-crosslinking experiments suggest that HoxTF is a heterodimeric DNA-binding activity and, in Neuro2a cells at least, is composed of 115kD and 90kD subunits as estimated by SDS-PAGE. Heterodimerisation is a common strategy for achieving differential gene regulation. Modified target-site specificity and altered regulatory capacity can be invoked by the association of particular partner-pairs or cofactors. For instance, the Hox/HOM genes themselves are able to bind to common target sequences sharing a TAAT core-motif (Gehring et al., 1994). Evidence also exists to suggest that cofactors can modify Hox/HOM target specificity, functioning by cooperative DNA-binding interactions (reviewed in Mann, 1995). For example, the Hoxb-1 protein is able to bind to, and activate transcription from, a consensus Hox binding-site (Pellerin et al., 1994). However, in conjunction with Pbx1 it binds to a different site containing the motif [T/G]GAT[T/G]GA[T/A]G (Pöpperl et al., 1995). Our EMSA experiments that examine the distribution of HoxTF indicate subtle discrepancies in the mobility of the protein-DNA complex in cell extracts of different origin. This may point to the existence of different heterodimeric states of the HoxTF binding activity in different cell-types.
Chapter 5

The G5 element: an additional cis-regulatory element in region C

5.1 5' Deletions identify a second regulatory element within region C

My experiments thus far have shown that efficient region C-directed transcriptional activation requires the presence of the HoxTF site located at bp 531-539. This binding-site is necessary for major aspects of region C enhancer function, such as expression in the mesoderm. Alone, it is capable of driving a limited pattern of expression in the central and peripheral nervous systems. I therefore wished to extend the analysis of region C to look for further regulatory elements. In particular ones that might cooperate with HoxTF to drive expression in the mesoderm and help in setting the anterior boundary of Hoxb-4 expression in the somites.

Initially I decided to analyse the sequences 5' of the HoxTF site. To facilitate the production of external deletion mutants of region C the parental reporter construct (WT) had been designed with suitable restriction sites for making nested deletions using exonuclease III (Henikoff, 1984). At an early stage of these studies a series of 5' and 3' region C nested deletions were made in the context of the WT reporter as described in the Materials and Methods Chapter (2.2.8). One of these mutants (Δ475) deleted the first 474bp of region C, removing all of exon 1 and the majority of the intronic sequences 5' of the HoxTF/YY1 site (see Figure 3.2). A second mutant construct was also made (Δ514) by digestion of the WT reporter with MuniI (see Materials and Methods Chapter, 2.3.1), deleting the sequences of region C to within 15bp 5' of the HoxTF site.

The two constructs were used to produce transgenic mice which were analysed as transient F₀ embryos and the results of these analyses are presented in Figure 5.1. The Δ475 construct showed a pattern of β-galactosidase activity that was indistinguishable from that of the WT construct (Figure 5.1a-c). Of the thirteen F₀ transgenic embryos generated, six expressed the transgene in a consistent manner. This indicates that the first 474bp of region C are not required for enhancer function. The Δ514 construct, however, showed a very different pattern of expression, similar to those seen with the HoxTF mutant constructs, MH/Y and MHTF2 (shown in Figures 4.7 and 4.8e-f, respectively). Of the nine F₀ transgenic embryos obtained with the Δ514 construct, six expressed the transgene in a consistent fashion. Staining due to β-galactosidase activity was present within the central and peripheral nervous systems of all expressing embryos. Neural tube staining showed the same distribution as
Figure 5.1: 5' Deletions identify a second regulatory element within region C

Another Hoxb-4 cis-regulatory element is located between bp 475-513 of region C. The β-galactosidase expression patterns observed in transient transgenic embryos carrying 5'-deletion constructs A475 and A514 are illustrated. A schematic of the reporter constructs and of the parental region C construct (WT) is shown below. The legend to the reporters is as described in Figure 3.3. The broken arrows denote the 5'-end of each construct relative to WT. The red cross just 3' of the MunI recognition sequence (M) shows the position of the HoxTF/YY1 binding-site. Exp. denotes the total number of positively stained embryos obtained that showed a consistent pattern of expression. Tg. denotes the total number of transgenic embryos.

(a) Shows a lateral view of a 11.0 dpc transgenic embryo carrying the WT region C construct for comparison.

(b) Lateral and (c) dorsal views of an 11.0 dpc transgenic embryo carrying construct A475. The pattern of β-galactosidase staining obtained is indistinguishable from that of the WT construct.

(d) Lateral and (e) dorsal views of an 11.5 dpc transgenic embryo carrying construct A514. Staining is visible within the neural tube (n) and elements of the PNS (p). This spatial pattern of neural staining mirrors that of the WT construct, with the same anterior limit in the neural tube at the spinal cord/hind brain boundary. The levels of expression are much lower, however. Some weak staining can also be seen in the lateral mesoderm posterior to the forelimb (m). There is also some posterior somitic staining in this embryo, although it is too weak to be visible.
those of the wild-type constructs (WT and Δ475), with an anterior limit at the spinal cord/hindbrain boundary. The levels of expression were very much lower, however (Figure 5.1d-e). Four out of the six embryos showed extremely weak staining in the somites posterior to the hindlimb. Five embryos exhibited staining in the lateral mesoderm between the limbs and again the levels of expression were much reduced compared to those of WT and Δ475. Also, anterior neural staining was visible in four out of the six positively stained embryos. These data indicate that a second cis-regulatory element is located between bp 475-513 of region C. The evidence suggests that this element is required for the proper levels of region C-directed expression within both neural and mesodermal tissues.

5.2 Identification of a G-rich binding-element in the bp 475-513 region

To identify the cis-regulatory sequences deleted by the Δ514 construct I performed analogous EMSA experiments to those used to identify the HoxTF/YY1 binding-site in CB1. An oligonucleotide probe was designed that overlapped the sequences deleted between constructs Δ475 and Δ514 by 13bp at the 5' end and 4bp at the 3' end (462-517). A probe covering bp 508-524 had previously shown no specific binding (Figure 4.3), suggesting that the sequences involved lay 5' of this. Oligonucleotide probes were labelled with [γ32P]-ATP, annealed, and purified by native polyacrylamide gel electrophoresis. EMSA experiments were carried out as described in the Materials and Methods Chapter (2.4) and the products were resolved on 5% (w/v) polyacrylamide gels. A schematic diagram summarising the results of these experiments is presented in Figure 5.2.

Specific DNA binding was observed with the 462-517 probe using whole-cell extracts from 8.5, 10.5 and 13.5 dpc total embryos and the mouse neuroblastoma cell-line Neuro2a. Further subdivisions of the bp 462-517 sequences located the binding activity, first to bp 462-501, and then to bp 475-501. This latter probe, however, showed weaker binding than 462-501, indicating that some additional 5' bases were required to stabilise the protein-DNA interaction. This hypothesis was confirmed by the probe 467-496 (WT) which behaved as 462-501 and detected one specific band (data not shown).

To further delimit the binding-site within bp 467-496 I systematically mutated the 470-493 sequences. Four probes were designed which carried contiguous 6bp mutations in the WT sequences (Figure 5.2, MA-D). The introduced mutations were of the least conservative variety, i.e. A to C, C to A, G to T and T to G transversions. These mutated probes were assayed for their ability to compete for the WT binding activity in EMSA experiments and the results are summarised in Figure 5.2. MA and MB behaved as WT whilst MC and MD showed no specific binding. These experiments revealed the presence of a single binding-site
A summary of the results of EMSA experiments used to localise the binding of factors within bp 475-513 of region C is illustrated. The *Hoxb-4* intron sequences encompassing bp 462-517 of region C (numbered above, upper strand only) are shown. The 5'-extents of constructs Δ475 and Δ514 are marked with broken arrows. Double-stranded oligonucleotide probes were derived from this sequence for use in EMSA experiments, the bounds of which are indicated by the oblongs below. Black denotes probes that detected no specific binding, whilst red denotes probes that specifically bound the Granville band. The numbers shown above define the limits of each probe and correspond to the region C sequence (Figure 3.2). Specific binding was found to be confined to the wild-type, 30bp, 467-496 probe (WT), the sequence of which is shown within it. The bp 470-493 sequences were systematically mutated in probes MA, MB, MC and MD. The mutated bases are underlined. MC and MD did not bind, but MA and MB did, defining the sequences CATGTGGGGAG (bp 482-493) as being partially or totally required (underlined above in black).
<table>
<thead>
<tr>
<th></th>
<th>462</th>
<th>475</th>
<th>501</th>
<th>517</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>AAGGGACCTCCGAGGCATGTGGGGGAGGGA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td>AAGTTCAAGCCGAGGCATGTGGGGGAGGGA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MB</td>
<td>AAGGGACCTAATCTTCATGTGGGGGAGGGA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MD</td>
<td>AAGGGACCTCCGAGGCATGTGTTTTCTGGA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
containing 5 contiguous G residues and localised to bp 482-493 of region C, where the sequence CATGTGGGGGAG is partially or totally required for binding. I named this binding-site G5 and the factor binding to it, Granville.

We wished to further characterise the G5 binding-site, in order to gain insight into the type of factor that Granville represents. Dr. Alejandro Gutman in the Rigby Laboratory has carried out detailed mutational analyses of the Granville binding-site, defining it as ATGTGGGGG. Single point mutations in any of the five contiguous G residues affect binding, although the 1st, 2nd and 5th G appear to be more important than the 3rd or 4th, and none of the mutations completely abolishes binding (Dr. A Gutman, personal communication).

5.3 In vivo analysis of a mutation in the G5 element

I wished to test whether the Granville binding-site, G5, had a cis-regulatory role in transgenic mice and if its specific ablation could reproduce the deletion phenotype of the Δ514 construct. To investigate this, a construct was generated that carried a mutation in the G5 binding-site of the Δ475 reporter construct (equivalent to WT, Figure 5.1). The cloning steps used to make this construct (Δ475MG) are described in the Materials and Methods Chapter (2.3.1). A 3bp mutation (CATGTGGGGGAG to CATGTTTTGGAG) was introduced into the G5 site that abolished Granville binding. When the wild-type probe (G5-WT) was incubated with an excess of cold oligonucleotide containing this mutation, no competition of binding activity was observed (Dr. A. Gutman, personal communication. Data not shown). The Δ475MG reporter construct was used to produce transgenic mice which were analysed as transient F₀ embryos at 10.5 dpc. Of the six transgenic embryos generated, two expressed the transgene in a consistent manner. The results of this analysis are presented in Figure 5.3. The observed expression pattern due to β-galactosidase activity obtained with Δ475MG was equivalent to that seen with the Δ514 deletion mutant previously described (Figure 5.1). Both embryos exhibited weak expression in the neural tube, posterior somitic mesodermal expression was observed in one of the positively stained embryos and lateral mesodermal expression in the other. These results indicate that the G5 binding-site is an essential cis-regulatory component of Hoxb-4. They also show that this site, like that for HoxTF, is necessary for proper tissue-specific transcriptional activation by region C within the mesoderm and nervous system. Unlike HoxTF, however, the G5 site does not seem to be an absolute requirement for mesodermal expression but it is required to achieve normal levels of enhancer activity.
Figure 5.3: The Granville binding-site is required for region C function but does not synergise with HoxTF

The β-galactosidase expression patterns observed in transient transgenic embryos carrying constructs Δ475MG and G+HTF are illustrated. A schematic of the reporter constructs and of the parental region C equivalent construct (Δ475) is shown below. The legend to the reporters is as described in Figure 3.3. The broken arrows denote the 5'-end of the Δ475 constructs relative to that of WT. The cross just 3' of the Muni site (M) shows the position of the HoxTF/YY1 binding-site. The circle immediately 5' of this denotes the wild-type (red) and mutant (black) G5 binding-sites. Exp. denotes the total number of positively stained embryos obtained that showed a consistent pattern of expression. Tg. denotes the total number of transgenic embryos.

(a) Shows a lateral view of an 11.0 dpc transgenic embryo carrying the Δ475 construct for comparison. The pattern of β-galactosidase staining obtained is indistinguishable from that of the WT construct (Figure 3.3).

(b) Lateral view of a 10.5 dpc transgenic embryo carrying construct Δ474MG. Weak staining can be seen in the neural tube (n) and somites (s) posterior to the hindlimb.

(c) Lateral and (d) dorsal views of an 11.5 dpc transgenic embryo carrying construct G+HTF, which incorporates the G5 and HoxTF/YY1 binding-sites of region C (bp 475-550). Expression is visible within elements of the PNS (p) but not within the CNS. This pattern of staining mirrors that of the single copy HoxTF site construct (1XHTF, Figure 4.10), shown for comparison in (e).
Δ475

bp 475

100bp

hsp 68

lacZ

SV40 pA

Exp. 6

Tg. 13

Δ475MG

bp 475

hsp 68

G+HTF

hsp 68

Exp. 2

Tg. 6

Exp. 9

Tg. 13
5.4 Granville and HoxTF do not cooperate to drive mesodermal expression

Studies described so far have successfully identified two cis-regulatory elements within region C. Namely, the HoxTF/YY1 site, binding HoxTF and YY1, and the G5 site binding Granville. Whilst the role of YY1 remains unclear, HoxTF is essential for region C-directed expression in mesodermal tissues and is required for correct neural expression. Granville, on the other hand, seems to be necessary for achieving proper levels of expression but not for tissue-specificity per se. An isolated HoxTF element is able to drive a limited amount of expression within the peripheral and central nervous systems. I next wanted to test whether HoxTF and Granville were able to cooperate to specify a more extensive expression pattern.

A construct was made that contained the G5 and HoxTF/YY1 sites cloned upstream of the hsp68-lacZ reporter gene. The cloning steps used to generate this construct (G+HTF) are described in the Materials and Methods Chapter (2.3.1). The fragment used contained the wild-type region C sequences from bp 475-550, maintaining the spacing of the two sites in their natural context. The G+HTF construct was used to produce transgenic mice which were analysed as transient F0 embryos at 11.5 dpc. The results of this analysis are presented in Figure 5.3c-d. Of the thirteen F0 transgenic embryos generated with the G+HTF construct, nine expressed the transgene in a consistent manner. The pattern observed was analogous to that seen for the HoxTF site alone (1XHTF). Expression was visible within elements of the PNS but not within the CNS, and no mesodermal staining was evident. These results suggest that whilst the G5 and HoxTF/YY1 sites are both required for the normal region C expression pattern in the paraxial mesoderm, the two elements together are unable to specify mesodermal expression. These data point to the presence of further cis-regulatory elements within region C, located 3' of the HoxTF/YY1 site.

5.5 In vitro analysis of the G5 and related binding-sites

I have shown that the G5 binding-site has an important cis-regulatory role in the function of the region C enhancer. I, in collaboration with Dr. Gutman, have also demonstrated that Granville represents a single binding activity in embryo and cell-line extracts, under the assay conditions used. Unlike HoxTF, Granville does not seem to be not an ubiquitous binding activity within the embryo. For example, Granville has not been detected in 10.5 dpc or 14.5 dpc liver extracts (Dr. A. Gutman, personal communication). In this respect it may be considered to have a tissue-specific pattern of expression whereas HoxTF does not.

Several vertebrate factors have been described that bind to G-rich target sequences (Faisst &
Meyer, 1992). Of the ones that have been cloned, most belong to the zinc finger family of transcription factors. Many zinc finger genes encode proteins that belong to the Cys$_2$-His$_2$ class and contain highly conserved 28-30 amino acid domains that tetrahedrally bind to a molecule of zinc through two cysteine and two histidine residues (reviewed in Berg, 1990; Rhodes & Klug, 1993). A large number of zinc finger proteins have been described which regulate transcriptional events during developmental processes. One particular group, encoded by the early growth response genes (EGR), belong to the family of immediate-early gene-encoded transcription factors. EGR-1 (NGFI-A, Krox-24), EGR-2 (Krox-20), EGR-3 and NGFI-C contain highly conserved zinc finger motifs and can bind to the same GCGGGGGCG target sequence (Swimhoff & Milbrandt, 1995 and references therein). The Wilms' tumour gene product (WT1) and the general transcription factor Sp1 possess related but less homologous zinc finger domains that also bind to related GC-rich sequences (Kadonaga et al., 1987; Hamilton et al., 1995; Nakagama et al., 1995). All of these factors were regarded as being possible candidates for the Granville binding activity, as was the helix-span-helix transcription factor AP-2 (Williams & Tjian, 1991). A comparison of the G5 and other G-rich binding-sites is shown later in this Chapter (Figure 5.5B).

Krox-20 (EGR-2) has been shown to be important for the regulation of the Hoxb-2 and Hoxa-2 genes (Sham et al., 1993; Nonchev et al., 1996). AP-2 is highly expressed throughout the neural crest lineage during embryogenesis (Mitchell et al., 1991) and, like Sp1, it is important for the regulation of a number of viral and cellular genes (Imagawa et al., 1987; Mitchell et al., 1987; Mitchell & Tjian, 1989; Nyborg & Dynan, 1990). Dr. Gutman has looked to see if the Granville binding activity is related to any of these known transcription factors by EMSA competition with known binding-sites for Krox-20, AP-2 and Sp1. With the G5-WT probe, two retarded bands are observed (data not shown); a lower, specific band corresponding to the Granville binding activity and, an upper non-specific band that is competed by a consensus Sp1 binding-site, indicating that it may be a result of non-specific Sp1 binding to G5-WT. Granville is not competed by a consensus AP-2 binding site, nor by a consensus Krox-20 site. However, a probe that constitutes a high-affinity binding-site for Krox-20 from the Hoxb-2 gene (Nonchev et al., 1996) does compete for Granville binding at a high concentration. A 200-fold molar excess of the Krox-20 oligonucleotide was equivalent in its ability to compete for the Granville band to a 20-fold molar excess of G5-WT itself (Dr. A. Gutman, personal communication). These data suggest that Granville may be related to the EGR family of zinc finger transcription factors but is probably not Krox-20.
5.6 Cis-regulatory elements required for mesodermal expression of Hoxb-4 are located within the left half of the intron.

I have determined that a fragment containing the the two region C cis-regulatory elements identified thus far (G5 and HoxTF) is insufficient to specify reporter gene expression within the somitic or lateral mesoderm (construct G+HTF, Figure 5.3). It is also clear that the first 474bp of region C are not necessary for enhancer function (construct Δ475, Figure 5.1). This suggests that one or more further elements, located 3' of bp 550 of region C (the 3' limit of the G+HTF construct), are required.

In order to identify further cis-regulatory elements I have begun 3'-deletional analyses of the Δ475 construct. A reporter was made which deleted the 3' 563bp of region C from Δ475. The cloning steps used to generate this construct (475-818) are described in the Materials and Methods Chapter (2.3.1). 475-818 was used to produce transgenic mice which were analysed as transient F_0 embryos at 11.0 dpc (Figure 5.4c-d). Of the twenty F_0 transgenic embryos generated, fifteen expressed the transgene in a consistent manner. 475-818 gave a pattern of expression that was equivalent to Δ475 and WT, although the expression levels seen were more variable. Also, a high proportion of these embryos (10/15) showed ectopic expression within the head region. These results indicate that the region C sequences from bp 475-818 contain the most important elements, in terms of a requirement for the tissue-specificities of enhancer activity, thus confirming the hypothesis that elements within sequences 3' of the HoxTF site are also required to drive expression within the mesoderm. Furthermore, it is possible to define these additional elements as lying between bp 551-818 of region C. The results also show that whilst the 3' half of region C is not required for most aspects of its expression pattern, the bp 819-1381 sequences seem to confer some degree of stability on the core-enhancer activity located in the 475-818 construct.

I also tested the ability of the deleted region, i.e. the bp 819-1381 sequences, to drive reporter gene expression (construct Δ819, Materials and Methods Chapter, 2.3.1). Eight F_0 transgenic embryos were obtained with this construct of which four showed consistent expression. At 11.0 dpc staining due to β-galactosidase activity was seen in the peripheral nervous system and lateral mesoderm but not in the somitic mesoderm (Figure 5.4e-f). The pattern derived from Δ819 bore a strong resemblance to that seen with the 1XHTF and G+HTF constructs (Figure 5.3). Examination of the region C sequences contained in this construct revealed the presence of an element between bp 940-948 that matches the HoxTF site consensus at 8/9 bases (GCCATTGGg). To test whether this element represented a second HoxTF/YY1 binding-site within region C, I analysed its binding characteristics in EMSA experiments. An oligonucleotide probe was designed which covered the bp 933-960 sequences of region C,
Figure 5.4: Cis-regulatory elements required for somitic *Hoxb-4* expression are located within the left half of the intron

The β-galactosidase expression patterns observed in transient transgenic embryos carrying constructs 475-818 and Δ819 are illustrated. A schematic of the reporter constructs and of the parental region C equivalent construct (Δ475) is shown below. The legend to the reporters is as described in Figure 3.3. Broken arrows denote the 5'-end of the Δ475 construct relative to that of WT. The relative positions of the G5 and HoxTF/YY1 binding-sites are shown as a red circle and cross respectively. Exp. denotes the total number of positively stained embryos obtained that showed a consistent pattern of expression. Tg. denotes the total number of transgenic embryos.

(a) Lateral and (b) dorsal views of of an 11.0 dpc transgenic embryo carrying the Δ475 construct for comparison (i.e. wild-type).

(c) Lateral and (d) dorsal views of an 11.0 dpc transgenic embryo carrying construct 475-818. The pattern of β-galactosidase staining obtained is comparable to that of the Δ475 construct. Note the expression within the somites (s).

(e) Lateral and (f) dorsal views of an 11.0 dpc transgenic embryo carrying construct Δ819. Expression is visible within elements of the peripheral nervous system (p) but not within the CNS. Additional staining is seen in the lateral mesoderm and in the head. This pattern of staining is similar to that of the 1XHTF and G+HTF constructs (Figure 5.3), suggesting the cis-regulatory involvement of a second HoxTF site.
encompassing this second putative HoxTF/YY1 site in the same relative position to that contained in the HoxC-WT probe, previously described in Chapter 4. This probe (HoxC-3’) was labelled with [γ³²P]-ATP and purified by native polyacrylamide gel electrophoresis. EMSA experiments were carried out as described in the Materials and Methods Chapter (2.4). However, purified HoxTF protein prepared by Dr. Marie Vandromme of the Division of Eukaryotic Molecular Genetics (NIMR) was used as a source of protein instead of whole-cell extracts. The HoxC-3’ probe is indeed able to bind HoxTF, although quite weakly. The HoxTF band is competed less efficiently by a range of molar excesses of the unlabelled HoxC-3’ probe, when compared to the 5’ Hoxb-4 intron site as a competitor (HoxC-WT; Figure 5.5A, lanes 1-7). A 50-fold molar excess of HoxC-3’ acts as a slightly more efficient competitor than a 10-fold molar excess of HoxC-WT (compare lanes 3 and 5 in Figure 5.5A). This indicates that the 3’ intron site is approximately three times weaker than the 5’ intron site with respect to HoxTF binding. I have also looked at the binding characteristics of HoxC-3’ in whole-cell extracts (data not shown). These investigations confirm that HoxC-3’ is also able to bind YY1 at a similar level to HoxC-WT.

Thus, preliminary studies indicate that region C contains a second HoxTF/YY1 site within the 3’ half of the intron and this may, in part, explain the expression pattern seen with Δ819. In summary, these results imply that the major elements required for region C tissue-specificity are located between bp 475-818. Within this fragment are two important binding-sites for HoxTF and Granville, that alone are only capable of specifying a limited pattern of neural expression. This demonstrates the presence of one or more cis-regulatory elements between bp 550-818 that can synergise with HoxTF and G5 to drive appropriate mesodermal and neural expression. Additional elements that could be required for the stability of the enhancer activity are located 3’ of bp 819 and may include a second HoxTF/YY1 site within region C, located between bp 940-948.

5.7 Summary and Discussion

This Chapter has described the mapping and characterisation of further cis-regulatory components of region C. Initially, 5’ deletions were employed to investigate the regulatory potential of sequences upstream of the HoxTF site. A deletion construct removing the first 474bp of region C (Δ475) was able to direct reporter gene expression in a manner indistinguishable from that of the wild-type construct. This shows that the deleted sequences, i.e. those of exon I and the initial 56bp of the intron, are not required for enhancer function. A further deletion removing the first 513bp of region C (Δ514) resulted in a marked decrease in the levels of enhancer activity, locating a second important regulatory element between bp 475-513. EMSA analyses were again employed to identify putative binding-sites within this
Figure 5.5: *In vitro* analysis of a second HoxTF/YY1 binding-site in region C

(A) Shows the results of an EMSA experiment revealing the presence of a second HoxTF/YY1 binding-site in region C, located within the 3′ half of the *Hoxb-4* intron. The oligonucleotide probe HoxC-3′ encompasses bp 933-960 of region C. MD1-WT (MD1) incorporates a high-affinity HoxTF site from *MyoD1* and has been described previously (see Figure 4.5). Probes were labelled with [γ³²P]-ATP as described in the Materials and Methods Chapter (2.4). Affinity-purified HoxTF (0.5μl per reaction, 5mg/ml total protein concentration) was used as a source of protein and was kindly provided by Dr. Marie Vandromme of the Division of Eukaryotic Molecular Genetics (NIMR). Specific competitor probes were added at a 10 to 250-fold molar excess as indicated above each lane. HoxC-WT represents the wild-type HoxTF/YY1 binding-site from region C (bp 524-550) previously described. MD1-M represents a mutant form of the *MyoD1* HoxTF site that does not bind HoxTF. The products of the binding reactions were resolved on a 5% (w/v) polyacrylamide gel. The position of the retarded band corresponding to HoxTF interaction is indicated. Free indicates the unbound probe.

(B) The top of the table shows a list of G-rich binding-sites for known transcription factors which bear strong similarities to the Granville binding-site G5. The early growth response transcription factors EGR-1, -2, -3 and NGFI-C all bind to similar sequences with the combined consensus shown (Swirnoff & Milbrandt, 1995). WT-1 also binds to the EGR sequence but that indicated has been determined as a high-affinity site (Nakagama *et al.*, 1995). The Sp1 binding-site ascribes to the consensus as described by Faisst & Meyer (1992) and the AP-2 site is from the SV40 enhancer (Williams & Tjian, 1991). The bottom of the table shows a sequence alignment of the region surrounding the G5 site from the *Hoxb-4* intron (G5) with a another potential Granville binding-site from 3′ half of the *Hoxb-4* intron (Gγ), and with element D from the mouse *Hoxa-4* gene (Galliot *et al.*, 1989). The area corresponding to the Granville binding-site as defined by our experiments (ATGTGGGGG) is boxed in white. Nucleotides which match the *Hoxb-4IG5* site and surrounding sequences are underlined.
**A**

<table>
<thead>
<tr>
<th>Probe Comp.</th>
<th>HoxC-3' WT1</th>
<th>MD1-WT</th>
<th>MD1-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>- 10 50 250</td>
<td>10 50 250</td>
<td>10 50 250</td>
<td>- -</td>
</tr>
</tbody>
</table>

**B**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Binding-Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGR</td>
<td>TGC\text{CTGGGCGG}</td>
</tr>
<tr>
<td>WT1</td>
<td>GCG\text{TGGGAGT}</td>
</tr>
<tr>
<td>Sp1</td>
<td>GGGGCGGGGG</td>
</tr>
<tr>
<td>AP-2</td>
<td>GCCTGGGGGA</td>
</tr>
<tr>
<td>Granville</td>
<td>ATGTGGGGGG</td>
</tr>
</tbody>
</table>

*G5 Sites from Hox b-4 and Hoxa-4*

- \text{Hoxb-4/G5}: TCCGAGGCCATGTTGGGAGGGGA
- \text{Hoxb-4/Gy}: TCC\text{TTTGGCCTG}
- \text{Hoxa-4/D}: GAGGC\text{GAGGGGAGGGGAGGGG**}
area and revealed the presence of a single G-rich binding-site. I have called this element G5 and the factor binding to it, Granville. By introducing a 3bp mutation in the G5 site of Δ475 (Δ475MG), that specifically abolished Granville binding, I was able to reproduce the Δ514 deletion phenotype. Transgenic embryos carrying Δ514 and Δ475MG exhibited the same tissue-specificities of expression as with the wild-type constructs, showing β-galactosidase activity in the central and peripheral nervous systems and in the mesoderm of transgenic embryos. Overall levels of activity of the transgenes were much reduced, however, and comparable to those seen with the HoxTF mutant constructs MH/Y and MHTF2 (Chapter 4). These data indicate that the G5 site acts to bolster the expression levels imposed by the region C enhancer but, unlike the HoxTF site, it is not required for tissue-specific expression.

The investigations thus far had maintained a minimalist approach to define specific regulatory sequences responsible for Hoxb-4 expression. An ultimate aim of this research is to reassemble the region C enhancer from its individual components, enabling the role of particular elements in the spatial restriction and tissue-specificity of Hox gene expression to be investigated. I tested whether the combined G5 and HoxTF sites were able to drive a more extensive pattern of reporter gene expression, when compared to the HoxTF site alone. A DNA fragment was tested that covered bp 475-550 of region C (G+HTF). This ensured that the natural genomic spacing of the two elements was conserved. Previous research has shown that in order for some transcription factors to interact effectively when bound to DNA, they must be stereospecifically aligned and, therefore, spacing may be an important factor (Takahashi et al., 1986). The pattern of expression derived from the G+HTF construct was essentially the same as that seen with the single copy HoxTF site (1XHTF). β-galactosidase staining was predominantly restricted to the peripheral nervous system and no somitic mesodermal activity was evident. These data indicate that Granville and HoxTF cannot synergise to drive expression outside of the pattern specified by HoxTF alone. Additional elements located within the 3' sequences of region C (bp 551-1381) must also be required. In light of previous observations that the G5 site is required for appropriate expression levels, it is interesting to note that those obtained from G+HTF were not significantly different from the 1XHTF construct. This suggests that Granville's effects on the levels of region C enhancer activity do not occur by a direct interaction with HoxTF, but via an indirect mechanism.

EMSA analyses were employed to further characterise the specificity of Granville binding. These experiments were not as straight forward as we had previously experienced with HoxTF due to a relatively weak shift seen with the G5 site (data not shown). It is not known whether this is because; G5 is a low-affinity binding-site for Granville, the Granville binding activity has a low relative abundance within our extracts, or the EMSA conditions used were not optimal for Granville binding. Initial studies defined the Granville binding-site as
CATGTGGGGGAG. Dr. Gutman in our laboratory has used single point mutations across this region to show that a core-motif of five contiguous G-residues is required for the efficient formation of protein-DNA complexes (Dr A. Gutman, personal communication). Mutational effects of the individual bases were not equivalent, with those within the 1st, 2nd and 5th G being more severe than those within the 3rd and 4th. Point mutations flanking the G5-core had no effect on Granville binding. However, Dr. Gutman has used combinatorial mutational analyses to further defined the G5 site as ATGTGGGGG.

A review of the current literature relating to known target sites suggested that Granville may belong to the zinc finger class of DNA binding proteins (Berg, 1990; Rhodes & Klug, 1993). Numerous transcription factors have been described that contain zinc finger motifs, many of which bind to GC-rich sequences (for review see Faisst & Meyer, 1992), some well characterised examples being Sp1 (Kadonaga et al., 1987) and members of the early growth response (EGR) genes. Krox-20 (Chavrier et al., 1988a, 1988b) is the mouse equivalent of human EGR-2 (Joseph et al., 1988; Chavrier et al., 1990) and belongs to a related family of zinc finger transcription factors that includes Krox-24 (also known as EGR-1, NGFI-A, Zif268 and Tis8), EGR-3 and NGFI-C (Christy et al., 1988; Lemaire et al., 1988; Sukhatme et al., 1988; Crosby et al., 1991; Patwardhan et al., 1991).

The EGR family of proteins possess highly homologous DNA binding domains composed of three Cys$_2$-His$_2$ zinc fingers. All can bind to the same target sequence (GCGGGGGCG) containing five contiguous G-residues, from which they can activate transcription (Lemaire et al., 1990; Crosby et al., 1991; Patwardhan et al., 1991). The Cys$_2$-His$_2$ zinc finger motif is one of the most widely occurring eukaryotic DNA-binding domain structures (Pellegrino & Berg, 1991). It consists of a conserved polypeptide domain of 28 to 30 amino acids that contains a number of invariant residues (Berg, 1990). The nuclear magnetic resonance solution structure of a single zinc finger has demonstrated that it forms a compact domain, stabilised by the coordination of a Zn$^{2+}$ ion through two cysteine and two histidine residues (Lee et al., 1989). The cocrystal structure of the three zinc fingers of Zif268 (EGR-1) bound to DNA has also been established by X-ray crystallographic techniques (Pavletich & Pabo, 1991). The zinc fingers are found to partly wrap around the DNA, following the major groove of the double-helix. Three amino acids in each finger interact with three bases within the major groove, thereby defining the specificity of binding. More recently Swimhoff & Milbrandt (1995) have sought to better define the binding-site preferences of individual members of the EGR family. This has enabled the derivation of a high-affinity combined consensus binding-site as TGCG[T/g][G/A]GG[C/a/t]G[G/T], where lower case letters indicate bases selected less frequently.
The Wilms' tumour gene product (WT1) and Sp1 possess zinc finger domains that are related but less homologous to those of EGR-1. WT-1 contains four zinc fingers, fingers II to IV share a high degree of homology with EGR-1 (Call et al., 1990), that can also bind to the same EGR target sequence. WT1, however, has been shown to act as a transcriptional repressor, playing a key role in development and differentiation of the urogenital system (Rauscher et al., 1990; Drummond et al., 1992). Sp1 is known to be involved in the transcriptional regulation of multiple genes, binding to GC-rich sequences that are important for the activity of several enhancers and promoters (Dynan & Tjian, 1983; Kadonaga et al., 1987).

In terms of identifying the Granville binding activity, members of the EGR family of proteins constituted strong candidates. Both Krox-20 and Krox-24 are expressed within the embryonic nervous system (Wilkinson et al., 1989a; McMahon et al., 1990), although the domain of Krox-20 expression is more anterior and does not overlap with that of Hoxb-4. High levels of early Krox-20 expression are confined to the presumptive regions of rhombomeres 3 and 5 (r3/r5) within the vertebrate hindbrain (Wilkinson et al., 1989a; Nieto et al., 1991), where it is required for their normal development (Schneider-Maunoury et al., 1993; Swiatek & Gridley, 1993). Krox-20 has been shown to act as a direct regulator of both Hoxa-2 and Hoxb-2 expression in r3/r5. However, isolated Krox-20 binding-sites are insufficient for r3/r5 regulation alone (Sham et al., 1993; Nonchev et al., 1996). EMSA competition experiments were used by Dr. Gutman to compare the G5 binding-site with previously characterised G-rich sites for Krox-20, Sp1 and AP-2. An oligonucleotide containing a single consensus Krox-20 binding-site was unable to compete for the binding of Granville. A high-affinity Krox-20 oligonucleotide containing two Krox-20 sites from the Hoxa-2 r3/5 enhancer (Nonchev et al., 1996), was able to partially compete for the Granville band but only at a high molar excess (200-fold). Taken in concert, these results indicate that whilst the G5 site may bind Krox-20 at some low level, the Granville binding activity does not seem to be Krox-20. It is possible that Granville is related to the EGR family of proteins and it may represent a novel member of this class of transcription factors. In light of this, continued work on the characterisation of Granville is being carried out by Dr. A. Gutman in our laboratory.

Unlike the HoxTF/YY1 site, G5 does not seem to be at all spatially conserved amongst the introns of Hoxb-4 genes or those of associated paralogues (data not shown). However, Galliot et al. (1989) have described the presence of six highly related G-rich sequences in the promoter of the mouse Hoxa-4 gene. These sequences were shown to be protected in DNaseI footprints by specific DNA-binding activities, present in both human HeLa and mouse F9 cell-lines. Comparison of the footprinted regions with the sequence of the G5 site reveals that they share many common residues, both in and around the G5 site (see Figure 5.5B).
Several of the footprinted regions were shown to contain genuine Sp1 binding-sites. However, an additional binding activity was also involved that migrated slightly faster than the Sp1-DNA complex in EMSA assays. This observation is consistent with our experiments on the G5 site. In the case of G5, an Sp1 oligonucleotide does specifically compete for the Granville band. Interestingly, however, the Sp1 site competes for a slower migrating non-specific band that migrates just above the Granville-DNA complex (Dr. A. Gutman, personal communication; data not shown). Subsequent experiments have shown that Krox-20 is capable of binding to the footprinted regions E and F of the Hoxa-4 promoter (Chavrier et al., 1990). These available data suggest that the G5 site may be related to the G-rich footprinted regions upstream of the Hoxa-4 gene and that they may bind similar factors, evoking a possible role for Granville-like factors in the regulation of two PG-4 Hox genes.

The final transgenic experiments described in Chapter 5 were aimed at defining further cis-regulatory elements located 3' of bp 550 of region C (the 3' limit of the G+HTF construct). Deletional analysis shows that the core-enhancer activity of region C is confined to a 344bp fragment located within the 5' half of the intron. A reporter construct carrying these sequences (475-818) was able to drive a wild-type pattern of expression. Levels of expression were greatly variable, however, and ectopic expression within the head was a feature of most embryos. These data brought to light two important points. Firstly, additional cis-regulatory elements are located between bp 551-818 and are required for region C enhancer activity in the mesoderm and appropriate activity in neural tissues. Secondly, they suggest that whilst regulatory components that are necessary to specify appropriate neural and mesodermal activity of region C are located between bp 475-818, other elements located in more 3' regions (i.e. between bp 819-1381) are required to stabilise this pattern. This information is summarised in Figure 5.6.

I also tested the ability of the bp 819-1381 sequences to drive reporter gene expression in transgenic mice. $\beta$-Galactosidase activity derived from this construct (A819) was present in a subset of the peripheral nervous system and lateral mesoderm and was extremely similar to that seen with the 1XHTF and G+HTF constructs. It is unlikely that the region C sequences that derive from the second exon of Hoxb-4 are involved in this regulatory capacity. The results of my 5'-deletional analyses show that the first 474bp of region C, including exon 1, are not required for region C function (Chapter 5.1). It seems highly improbable that sequences from exon 2 (bp 1219-1381) that encode the functionally conserved homeodomain, also have a role in cis-regulation. Examination of the bp 819-1381 sequences of region C revealed the presence of an additional HoxTF/YY1 site (located between bp 940-948) that is capable of HoxTF and YY1 binding in vitro (Figure 5.5A). However, the effect of specifically mutating this second HoxTF/YY1 element in the transgenic assay has not yet been tested. I have also identified a sequence within the 3' half of the intron that displays
extended similarity to the G5 site but is in the opposite orientation (located between bp 1014-1006 on the reverse strand, see Figures 3.2 and 5.5B). These data imply that there is at least a functional HoxTF/YY1 site, and possibly a G5 site in close proximity, present within the 3’ half of the Hoxb-4 intron. The evidence for this is mainly circumstantial. However, the similarities between the expression patterns derived from the 1XHTF/G+HTF and the Δ819 constructs are extremely provocative.

A summation of my observations enables a possible mechanistic model to be put forward for the interaction between regulatory components located in the Hoxb-4 intron and the associated promoter (Figure 5.7). The most favoured theory for the mechanism of enhancer action involves the DNA-looping model (Ptashne, 1986, 1988). This model assumes that enhancer-driven promoter activation is a consequence of direct protein-protein interaction between promoter- and enhancer-bound factors, accompanied by looping of the intervening DNA sequences. DNA looping has been most studied in prokaryotic systems where it was initially proposed to explain the interaction of transcriptional regulators bound to spatially separated DNA sites (Dunn et al., 1984; Majumdar & Adhya, 1984). More recently in eukaryotic systems, looping interactions between homologous factors have been visualised by electron microscopy for the progesterone receptor and for Sp1 (Theveny et al., 1987; Mastrangelo et al., 1991; Su et al., 1991), and for heterologous proteins between Sp1 and the bovine papillomavirus enhancer protein, E2 (Li et al., 1991). The self association of DNA-bound Sp1, bringing together distant DNA segments, is dependent on the presence of a glutamine-rich trans-activation domain. Presumably, other related factors which also possess glutamine-rich regions, such as Granville which may be related to EGR/Sp1, could also function in a similar manner.

In the case of region C, protein-protein interactions between factors binding to the two G5 sites would enable the formation of a loop. This would bring the sequences binding HoxTF, and the associated factors, into close proximity with one another (to within 8bp). The nucleoprotein complex could presumably then loop back on itself to drive transcription from the promoter. In the case of region C interacting with the Hoxb-4 promoter, this could involve a trimeric association of the three HoxTF elements enabling interaction between regulatory elements in the intron and promoter. In support of this, I have previously noted that the role of elements within the bp 819-1381 sequences seems to be as a stabiliser of the core enhancer activity, located between bp 475-818. I have also observed that three multimerised copies of the HoxTF binding-site offer a much stronger and more stable pattern of expression than a single one alone.

In summary, I have shown in this Chapter that the major elements required for region C-directed tissue-specific expression are located within the intron between bp 475-818. These
include a binding-site for HoxTF/YY1 and additional *cis*-regulatory elements located between bp 551-818 that are as yet unknown. Appropriate levels of enhancer activity require the presence of the G5 element, located between bp 483-491. This site is a target for a factor that we have named Granville, potentially related to zinc finger proteins of the EGR-1/Sp-1 group. An additional HoxTF/YY1 site and possible Granville site are located within the 3’ half of the intron and are thought to be required for stable enhancer activity, for which I have suggested a mechanistic model involving DNA looping.
Hoxb-4 region C

Enhancer Activity

Meso/CNS/PNS

LM/PNS

Stable Enhancer Activity
Figure 5.6: Regulatory components of the *Hoxb-4* region C enhancer

A schematic representation showing the region C enhancer and the transcriptional regulatory activities that lie within it, as defined by this research, is illustrated. The legend to region C is as described in Figure 3.3 and used throughout this study. The relative positions of binding-sites for Granville (green-gold hexagonals) and HoxTF/YY1 (blue/purple pentagonals) within the intron are shown. Major restriction sites are indicated; A (Asp718), B (BglIII), M (MunI), S (SalI), Sf (SflI).

The core-enhancer activity of region C is confined to a 344bp fragment located within the 5'-half of the intron, containing functional G5 and HoxTF sites. This lies between bp 475-818 and is able to drive a wild-type pattern of expression in the mesoderm (Meso), central nervous system (CNS) and peripheral nervous system (PNS), but with variable levels of expression. A second domain of enhancer activity able to drive expression in the lateral mesoderm (LM) and peripheral nervous system (PNS) is located in the 3'-half of region C between bp 819-1381. This region also contains a HoxTF/YY1 binding-site and a putative G5 site. When combined, the two enhancer subdomains are able to drive stable region C-like expression.
A possible mechanistic model for the interaction of factors bound to region C cis-regulatory elements by the formation of a looped structure is presented. A schematic representation of the *Hoxb-4* intron is illustrated above (broken black line) and is not to scale. The sequence and relative positions of the Granville (G5' and G3') and HoxTF/YY1 (H5' and H3') binding-sites are shown, with their relative orientations indicated by arrows.

In this model, shown below, protein-protein interaction between factors binding to the two G5 sites (G5' and G3') would enable the formation of a loop, bringing the two HoxTF sites (H5' and H3') into close proximity of one another, potentially enabling direct HoxTF-HoxTF interaction. The relative positions of the conserved motifs (I-IV) of CB1 within the intron are also shown. Presumably the nucleoprotein complex formed could loop-back on itself to modulate transcription from the associated promoter.
Final Discussion

Studies of the mechanisms of vertebrate gene regulation are confronted with a common problem. How can you sift through noncoding sequences, which in mammals represents some 95% of the genome, to uncover the regulatory information that lies buried within them? Transgenesis provides a suitable *in vivo* assay for testing functionality but is arduous. Therefore, any technique that can streamline this process is extremely beneficial. My starting point for the definition of *cis*-regulatory elements within region C involved comparative sequence analysis to look for conserved regions of the *Hoxb-4* intron. The hypothesis adopted was that at least some elements important for *Hox* gene regulation, particularly those with a fundamental role, may have been conserved throughout evolution.

This approach was successful in that I identified a block of sequence homology (CB1) conserved amongst Paralogous Group 4 *Hox* gene introns from mouse, chicken and *Fugu*. CB1 proved to be functionally important for region C activity in the transgenic assay, thereby verifying the use of this strategy. At a finer level of resolution, however, sequence comparison has been less informative. Deletion of the most conserved sequences of CB1 from region C had no apparent effect in the transgenic assay. Furthermore, the *cis*-regulatory sequences that I have identified as being important for mouse *Hoxb-4* regulation in this study are either relatively unconserved, e.g. the HoxTF/YY1 site in CB1, or not conserved at all, e.g. the G5 site. These results raise some important reservations about the validity of sequence comparison as a definitive tool for such analyses. From a practical standpoint the question that needs to be asked is: Does sequence comparison lead to a more rapid identification of *cis*-regulatory elements than a transgenic approach alone? From my experience and with the benefit of hindsight, the answer to this question inclines towards the negative. It is true that the identification of CB1 lead to the discovery of the HoxTF/YY1 binding-site more quickly than might otherwise have occurred. On the other hand, the conservation of sequences within CB1 actually proved to be quite misleading as I was unable to assign any function to them.

Gérard *et al.* (1993) were the first to investigate the role of defined conserved domains with respect to *Hox* gene regulation using the comparative sequence analysis-transgenesis strategy. In their work, comparison of *Hoxd-11* sequences encompassing the mouse and chicken genes revealed the presence of seven areas of significant homology between noncoding regions. However, only some of these conserved regions could be shown to have a regulatory role in
the transgenic assay. In an attempt to refine this approach we sought to use Fugu as an experimental system for interspecies sequence comparison. Fugu has the smallest known vertebrate genome due to an average relative reduction of about 7.5-fold in the amount of noncoding DNA, including unexpanded introns. For instance, dystrophin is the the largest known mammalian gene spanning some 2.5Mb, whereas the Fugu equivalent is thought to be approximately 10-fold smaller despite retaining a similar gene structure (Elgar et al., 1996). Thus, it was hoped that the Fugu Hoxb-4 intron would represent a functionally compact domain, greatly suited for sequence comparison against region C. Somewhat surprisingly, the Fugu Hoxb-4 intron transpired to be approximately 85% larger than that of the mouse (1.5kb compared 0.8kb). On the other hand, the chicken intron is slightly smaller (0.7kb). In my experience, the use of Fugu for sequence comparison has been no more effective than any other vertebrate species may have been. For instance, mouse-Fugu Hoxb-4 intron sequences generated no more homology information about CB1 than did the mouse-chicken comparison. Furthermore, the Fugu intron appears to be non-functional in that it is insufficient to drive reporter gene expression in transgenic mice, whereas the chicken intron contains a region C-like activity (Morrison et al., 1995).

The discrepancy in size for the mouse and Fugu Hoxb-4 introns is intriguing but is not a peculiarity since the Fugu Hox clusters are of comparable size to those in other vertebrates (see Krumlauf, 1994; Elgar et al., 1996). This observation further strengthens the presumed importance of structural constraints on the function of the Hox cluster throughout evolution. The importance of Pc-G and trx-G proteins for latter phases of Hox/HOM gene expression has been discussed at length in the Introduction. These factors are thought to act by the modification of chromatin structure and it is possible that the size of the Hox cluster has been at least partly maintained as a result of this. Thus, the chromatin state of one gene need not impinge upon that of a neighbour if there is sufficient space between them. It is also thought that the Hox clusters have remained as a unit due to a regulatory requirement for precise spatiotemporal gene activation (Duboule, 1992). It may well be the case, therefore, that the structural conservation of the Hox complexes is maintained due to evolutionary pressure from the integral effects of these two aspects of regulation.

We have also tried to use Fugu sequence comparisons in our laboratory to examine the regulation of the myogenic regulatory factors Myf-5 and MRF-4 (Herculin). In the mouse these two genes are orientated in the same transcriptional sense and are closely linked, being separated by an intergenic region of approximately 7.2kb. The assumption is that this linkage is functionally important, possibly due to the presence of shared regulatory elements as in the Hox clusters. Data derived from transgenic analyses of the mouse Myf-5 gene have indicated that its regulation involves the cooperation of multiple upstream and downstream regulatory elements. However, approximately 15kb of sequences flanking Myf-5 and including a
significant amount of MRF-4 are insufficient to recapitulate the full expression pattern of Myf-5. Oliver Coutelle in our laboratory has cloned and sequenced an equivalent genomic region from Fugu. Somewhat significantly, perhaps, Myf-5 and MRF-4 also linked in this species, further reinforcing the possibility of shared elements. The Fugu intergenic region is approximately half the size of that of the mouse, but despite the conservation at the genetic level, no significant sequence homology exists between noncoding regions of the mouse and Fugu loci. Where Fugu may be useful is in multispecies sequence comparisons between regions of strong conservation, thereby providing an additional level of stringency. Indeed, three way comparisons between mouse, chicken and Fugu have aided Pöpperl et al. (1995) to quickly identify conserved elements that are required for Hoxb-1 autoregulated expression in rhombomere 4. The caveat to these experiments is that the three tested regions were shown to be functionally equivalent in the transgenic assay, therefore lending a greater degree of confidence to the results of the sequence comparison. It would seem on the balance, therefore, that sequence comparison is a potentially useful but somewhat unreliable tool for defining cis-regulatory sequences.

The research I have conducted for this thesis has proved to be successful in that I have achieved most of my initial goals. I have identified at least two important cis-regulatory elements of Hoxb-4; namely the HoxTF/YY1 binding-site at bp 531-539 of region C, and the Granville binding-site (G5) at bp 483-491. The general location of other important elements within the intron may also be inferred. These include; elements that are essential for mesodermal expression located in the 5' half of the intron, and the possibility of additional HoxTF/YY1 and Granville binding-sites located in the 3' half of the intron that are required for stable enhancer activity. Based on this work, there is much scope for future research and I would finally like to discuss some of the key issues, together with possible approaches.

The continued identification of other cis-regulatory elements within region C should be relatively straightforward. In terms of the elements required for mesodermal regulation, these have been confined to a 344bp fragment lying between bp 551-818. Further definition ought to be quickly achieved by the sequential addition of 3' sequences to the G5+HTF construct (bp 475-550), until mesodermal expression is restored. This strategy, rather than the contrasting one of making 3' deletions from the 475-818 construct, will enable the sufficiency of the CB1 sequences to be tested. This is a complementary question to that asked by the deletion of the most conserved sequences of CB1 (Δ557-597), which showed no apparent effect on transgene expression compared to that of WT. It will also be important to look at the significance of the second HoxTF/YY1 site located between bp 940-948, and of the potential Granville binding-site at bp 1014-1006. Mutation of these sites in the context of region C will enable my model of DNA-looping to be investigated. Deletion of these two putative elements in construct 475-818 has invoked a role for them in stabilising the enhancer
activity. It is not clear to me quite what the predicted expression pattern would be from mutating these sites, and it may be less or more severe than that of the deletion.

Throughout my transgenic experiments I have only analysed region C mutations on the hsp68 promoter. It is now important to look at the effects of mutating the defined elements in the context of the Hoxb-4 gene. This will be extremely informative for several reasons; firstly it will enable a more accurate assessment of the effects of the mutations in terms of spatial and temporal Hoxb-4 regulation. Secondly, it will be possible to look at the interactive and competitive roles of cis-regulatory elements within the promoter and enhancer, or between enhancers. Apparently, enhancer-promoter element interactions are important for obtaining the proper relative distribution of Hoxb-4 expression along the AP axis, since on the hsp68 promoter region C cannot generate sharp anterior boundaries. Thirdly, the effects of mutations on the expression of other genes within the Hoxb cluster can be tested by a gene targeting approach. This is of particular interest because the anterior boundary of neural tube expression generated by region C is that of the next 5' gene along the complex, Hoxb-5 (Whiting et al., 1991). Thus, it will be possible to ascertain whether region C is required for the normal expression of Hoxb-5 and may shed some light on the requirements for the structural organisation of the Hox cluster.

A final and most important area of research which is currently underway relates to the further characterisation and cloning of HoxTF and Granville. I have attempted to clone HoxTF using the two direct approaches, i.e. south-western screening of an F9 EC cDNA phage library using a binding-site probe, and genetic complementation in yeast using the one-hybrid system with embryonic libraries. Although our crosslinking experiments have indicated that HoxTF binds to DNA as a heterodimer, this does not necessarily mean that either of the partners in this interaction cannot bind in the absence of the other. For this reason it was worth trying the direct cloning methods, which unfortunately have proved unsuccessful. Of course, this is not the only reason why these strategies may not have worked. The DNA binding form of HoxTF may require some form of post-translational modification not present in bacteria and possibly absent in yeast. Another possibility is that for either HoxTF partner the DNA binding domain may lie at the 5' end of the gene. If HoxTF transcripts are large then the abundance of functional clones within a given oligo-dT primed library may be small. For this reason, the yeast one-hybrid screen was attempted using both oligo-dT and random primed libraries. Dr. Marie Vandromme in our laboratory has now begun purify to HoxTF, with a view to indirect cloning after peptide sequences have been determined. We are extremely hopeful of the success of this approach, particularly as the purification strategy has proven to be relatively straightforward. We are fortunate in this respect in that the HoxTF site from MyoDI represents a high-affinity binding-site, well suited to an oligonucleotide binding-site affinity strategy. In addition, the DNA-binding form of HoxTF from F9 EC cells
is apparently glycosylated and can be greatly enriched using wheat-germ agglutinin affinity-chromatography. Granville, on the other hand, may be more problematical with regard to cloning for two main reasons. Firstly, the G5 binding-site that we have defined appears to be relatively weak, making any available cloning strategy more difficult. This could be due to G5 being a low-affinity site, Granville being a low-abundance factor, sub-optimal assay conditions, or a combination of these factors. Secondly, since Granville appears to be a zinc finger protein related to the EGR family and binds to a G-rich sequence, the potential exists for a high rate of false positive detection during cloning. Therefore, further experiments are required to clarify such issues and are being investigated by Dr. Gutman in our laboratory.

The role of HoxTF in the regulation of *Hox* genes appears to be significant given the requirement for HoxTF binding-sites for the activity of both the *Hoxb-4* promoter and region C. The presence of HoxTF binding-sites in the 5' regions of the myogenic regulatory factor genes, *myogenin* and *MyoD1*, is extremely tantalising. It is possible that HoxTF represents an upstream regulator of both of these sets of important developmentally regulated transcription factors, both of which are expressed within the mesoderm. In support of this we have also identified a HoxTF binding-site in the promoter region of the *Myf-5* gene (Oliver Coutelle, personal communication). The potential role of HoxTF in myogenic factor regulation should be relatively straightforward to test. Dr Sui-Pok Yee, previously a member of our laboratory, has shown that complete recapitulation of the *myogenin* expression pattern can be achieved using 133bp of 5'-flanking DNA in transgenic mice (Yee & Rigby, 1993). The HoxTF site lies in this region and the effect of mutating it can be easily tested. Another point which requires clarification is the involvement of YY1 binding to the HoxTF/YY1 sites. The result of mutating the intron YY1 binding-site in the MYY1 construct was ambiguous, since the observed effect of a reduction in the levels of mesodermal expression was potentially attributable to a reduction in HoxTF binding-affinity (see Chapter 4). This could be tested by switching the intron HoxTF site for a stronger YY1 mutant version, such as that from *myogenin* that does not bind YY1, to see if it is possible to override the effect. Now that we have access to purified HoxTF, we can also investigate its affinity for a range of wild-type and mutant oligonucleotides *in vitro* in both the presence and absence of YY1 protein. This should illuminate whether HoxTF and YY1 binding to overlapping sites is mutually conducive or exclusive.
Bibliography

Acampora, D., D'Esposito, M., Faiella, A., Pannese, M., Migliaccio, E., Morelli, F., Stornaiuolo, A., Nigro, V.,

Affolter, M., Percival Smith, A., Muller, M., Leupin, W. & Gehring, W. J. (1990). DNA binding properties of


166


176


---

185


192


Multiple Positive and Negative Regulatory Elements in the Promoter of the Mouse Homeobox Gene Hoxb-4

ALEJANDRO GUTMAN, JONATHAN GILTHORPE, AND PETER W. J. RIGBY*
Laboratory of Eukaryotic Molecular Genetics, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, England

Received 2 June 1994/Returned for modification 6 July 1994/Accepted 25 August 1994

Mouse Hoxb-4 (Hox-2.6) is a homeobox gene that belongs to a family which also includes Hoxa-4, Hoxc-4, and Hoxd-4 and that is related to the Deformed gene in Drosophila melanogaster. We have determined the sequence of 1.2 kb of 5′ flanking DNA of mouse Hoxb-4 and by nuclease S1 and primer extension experiments identified two transcription start sites, P1 and P2, 285 and 207 nucleotides upstream of the ATG initiator codon, respectively. We have shown that this region harbors two independent promoters which drive CAT expression in several different cell lines with various efficiencies, suggesting that they are subject to cell-type-specific regulation. Through detailed mutational analysis, we have identified several cis-regulatory elements, located upstream and downstream of the transcription start sites. They include two cell-type-specific negative regulatory elements, which are more active in F9 embryonal carcinoma cells than in neuroblastoma cells (regions a and d at −226 to −186 and +169 to +205, respectively). An additional negative regulatory element has been delimited (region b between +22 and +113). Positive regulation is achieved by binding of HoxTF, a previously unknown factor, to the sequence GCCATTTG (−148 to +155) that is essential for efficient Hoxb-4 expression. We have also defined the minimal promoter sequences and found that they include two 12-bp initiator elements centered around each transcription start site. The complex architecture of the Hoxb-4 promoter provides the framework for fine-tuned transcriptional regulation during embryonic development.

Homeobox genes encode a large family of transcription factors and are found in a wide range of species, both invertebrate and vertebrate (reviewed in references 12, 17, and 23). In Drosophila melanogaster, the genes of the Antennapedia/Ultrabithorax complexes, which are clustered together, specify the phenotypes of individual segments along the anteroposterior axis of the body. In mammals, there are four clusters of Antennapedia-class homeobox genes, called Hox genes, each on a different chromosome, which are presumed to have arisen by duplication and expansion of an ancestral complex. The genes in each of the four clusters that derive from the same ancestral gene are called paralogs. One of the most remarkable features of both the invertebrate and the vertebrate Hox clusters is that there is a direct relationship between the position of a gene in the cluster and its expression domain. The genes at the 5′ end, in the transcriptional sense, have the most posterior domains, and the genes at the 3′ end have the most anterior.

The role of Hox genes in mammalian development has been demonstrated by both ectopic expression (3, 16, 22, 39) and by the introduction of loss-of-function mutations (5, 6, 18, 21, 27). These data make it clear that the proteins encoded by these genes are involved, as in D. melanogaster, in the specification of structures along the anteroposterior axis. Thus, if a gene is ectopically expressed in regions more anterior than its normal expression domain, transformations to more posterior structures are observed, while if a gene is inactivated, it can lead to anterior transformations. It is therefore clear that the exact setting of the boundaries of expression of the Hox genes along the anteroposterior axis is crucial to orderly development.

We, and others, have therefore sought to identify the DNA sequences that control transcription of Hox genes during development. Experiments using transgenic animals have shown that it is possible to recapitulate the normal expression pattern. In two cases, Hoxa-7 and Hoxb-4, multiple regulatory elements located downstream of the transcription start site are required (26, 38). However, these elements operate in opposite fashions. In the case of Hoxa-7, the promoter, the 5′ flanking DNA, is active in all regions of the embryo, and this activity is restrained by the regulatory elements (26). In the case of Hoxb-4, the 5′ flanking DNA is essentially inactive, and proper expression is seen only in the presence of the regulatory elements. We have, moreover, shown that the regulatory elements of Hoxb-4 act as spatially specific enhancers in that they are capable of imposing correct boundaries of expression along the anteroposterior axis on heterologous promoters (38).

Although the 5′ flanking DNA of Hoxb-4 is inactive in the mid-gestation embryo in the absence of the enhancers, it is clear that the promoter must contain elements involved in responding to them and in achieving proper levels of transcription. In order to characterize such elements, we have investigated the transcriptional regulation of the gene in cultured cells in which the 5′ flanking DNA is active. These experiments have enabled us to define the architecture of the promoter and to identify, in addition to several negative regulatory elements, a binding site for a previously unknown activator which is essential for efficient transcription of the gene.

MATERIALS AND METHODS

DNA sequencing. The nucleotide sequence of both strands of the DNA was determined, at least twice, by the dideoxynucleotide method (29), using 32P-dATP.

Recombinants. The −926/+285 reporter recombinant was constructed by inserting a PstI-AseI fragment (the AseI site was blunted with Klenow) of the mouse Hoxb-4 promoter into the PstI-BglII restriction sites (the latter made blunt by filling in with Klenow) of the pBLCAT3 vector (20).

The +285/−926 construct was made by cloning this same fragment into the
were created by cutting the -926/+285 reporter with blunt or constructed from -926/+285 by deleting a fragment delimited by s1 nuclease III, S1 nuclease, T4 polynucleotide kinase, and T4 DNA ligase (the precise boundaries of the clone were determined by sequencing). An oligonucleotide was used to replace the HindIII-SacII fragment of -926/+285 by the -32 to +21 sequences, giving -32/+285. +21/+285 mut was constructed by inserting a MnlI linker into the single PmlI site of +21/+285. Another oligonucleotide was inserted into the Mscl (partial digestion)-BglII sites of -926/+285, generating -926/+205. -926/+169, -926/+149, and -926/+113 were created by deleting the -926/+285 reporter with BglII and PmlI, Mscl or BspMl, respectively, followed by filling in and religation. -186/+205 and +21/+205 were derived from -926/+205 by deletion of a HindIII-apal or a HindIII-SacII fragment, respectively. A similar strategy was used for creating -186/+169 and +21/+169 from -926/+169, -186/+149, and -186/+113 from -926/+113. Oligonucleotides with HindIII-BglII ends were used to construct -32/+21, +46/+86, +113, +21, and +75/+86. Mutations in the Spl-like binding sites of -32/+21 and +46/+86 were introduced by cloning the following oligonucleotides in pBLCAT3 (mutated bases are underlined; compare with the wild-type sequence in Fig. 1).

HSPCAT was constructed by inserting the 0.3-kb SalI-BamHI fragment of the heat shock gene hsp86 promoter into the same sites of pBLCAT3. pBLCAT2 and pBLCAT3 vectors have been previously described (20). A double-stranded oligonucleotide containing the +135 to +169 sequences of Hoxb-4 (Fig. 1), with added HindIII ends, was cloned in both orientations into the HindIII sites of the following recombinants: B-CAT (19), pBLCAT2 (20), -5/+7 (see Fig. 3), and -32/+21 (see Fig. 3), producing ACT-BAT-Catd and ACT-BAT-Catr, ACT-PBLCAT2d and ACTPBLCAT2r, ACT-Ind and ACT-Indr, and LACT-Mpd and LACT-Mpr, respectively. Another double-stranded oligonucleotide containing the +143 to +169 sequences of Hoxb-4 (Fig. 1) was cloned into the HindIII site of -32/+21, producing in the direct and reverse orientations ACT-WT and ACT-WTr, respectively. HoxTFmut, YY1mut, and Double mut were obtained by cloning the M8, M5, and M3 double-stranded oligonucleotides (see Fig. 7) in the HindIII site of -32/+21 (previously filled in with Klenow).

Transfections. Subconfluent 9-cm-diameter plates were transfected by the calcium phosphate method (14) with 10 to 15 μg of a reporter recombinant, 5 μg of pBluescript KS+ and 4 μg of the internal control plasmid RSV-Bgal (13) or 0.25 to 1 μg of pBCTBX2 (15) in S1 nuclease experiments. Cells were incubated with the calcium phosphate precipitate in culture medium containing 10% (vol/vol) fetal calf serum (FCS) for 20 h and then washed and grown for an additional 28 h in complete medium. Each experimental point was determined in duplicate or triplicate in at least three independent transfections. We have calculated the standard deviations on all of our experiments. These are generally between 10 and 20% of the stated value and never exceed 25%. β-Galactosidase and chloramphenicol acetyltransferase (CAT) activities were measured as described previously (28, 34). Quantitation of CAT activity was performed with a phosphomager or by scintillation counting. F9 embryonal carcinoma (EC) stem cells were differentiated by incubation for 3 days in medium supplemented with 20% (vol/vol) fetal calf serum, 1 mM cyclic AMP (cAMP), 0.1 mM isobutyl methylxanthine and 5 x 10^-8 M retinoic acid. PC12 cells were first grown in suspension and then attached to collagen-coated plates for 2 weeks, when they were differentiated with 50 ng of mouse 2.5S nerve growth factor (Sigma) per ml for 10 days.

S1 nuclease analysis. A SacI-EcoRI DNA restriction fragment encompassing nucleotides -211 to +151 (Fig. 1) of mouse Hoxb-4 was dephosphorylated with 0.1 U of calf intestinal phosphatase and then heated at 85°C for 30 min to inactivate the enzyme and labelled with [γ-32P]ATP. The labelled fragment was purified by a minicolumn (Magic Clean-up System, Promega Inc.) and denatured in 50% (vol/vol) dimethyl sulfoxide-0.5× TBE (0.045 M Tris-borate, 0.001 M EDTA [pH 7.5]), and both strands were separated in a 5% (wt/vol) polyacrylamide gel (60:1, acrylamide-bisacrylamide), in 0.5× TBE and run at 8 V/cm. An excess of probe (10^6 cpm; 10 fmol) was hybridized overnight at 52°C with 150 μg of total RNA in 30 μl of a solution of 0.04 M pipervine-N^1,N^7-bis-2-ethanesulfo-
IONIC ACID (PIPPES; pH 7), 0.4 M NaCl, and 1 mM EDTA. Digestion with S1 nuclease was performed at 37°C, as described by Sambrook et al. (28). The reaction products were resolved in a 7 M urea-6% (wt/vol) polyacrylamide gel. To quantify CAT mRNA in transfected cells (see Fig. 6B), an EcoRI-HindIII restriction fragment from pBLCAT3 was employed under the same conditions described above. In addition, an 80-mer oligonucleotide labelled with [γ-32P]ATP was used as a probe to detect mRNA from the internal control plasmid pBCTBX2 (15).

Primer extension. An oligonucleotide primer (+115/+179) [Fig. 1] was labelled with [γ-32P]ATP (20,000 cpm/μl; 3 x 10^6 cpm/μmol) and incubated with 100 μg of total RNA in 30 μl of a solution of 0.04 M PIPES (pH 7), 0.4 M NaCl, 1 mM EDTA, and 80% (vol/vol) formamide for 10 min at 85°C. This mixture was then incubated overnight at 30°C and treated with reverse transcriptase, as previously described (28). The reaction products were resolved in a 7 M urea-6% (wt/vol) polyacrylamide gel.

DNA mobility shift assays. Oligonucleotides were labelled with [γ-32P]ATP, annealed, purified by native polyacrylamide gel electrophoresis, and used as probes in gel retardation experiments (1,000 cpm per fmol, and 10 fmol per reaction mixture). Each reaction mixture contained, in a 20-μl volume, 10 to 20 μg of whole cell extract, 1 μg of calf thymus DNA, a specific competitor as indicated, 50 mM Tris-HCl (pH 7.9), 6 M MgCl2, 0.2 mM EDTA, 1 mM dithiothreitol, and 15% (vol/vol) glycerol and was incubated at 0°C for 15 min. The probe was then added, and incubation proceeded for 10 additional min at 30°C. Afterwards, the samples were loaded onto a 5% (wt/vol) polyacrylamide gel, in 0.25X TBE, and electrophoresed at 150 V for 2 h. The gel was then dried and autoradiographed.

Nucleotide sequence accession number. The DNA sequence presented here has been deposited in the EMBL database under accession number X71912.

RESULTS

Sequence of 1.2 kb of Hoxb-4 5' flanking DNA and identification of the transcription start sites. Our transgenic experiments showed that the recapitulation of the expression pattern of Hoxb-4 in the embryo requires only 1.2 kb of DNA 5' to the translation initiation codon (38). We therefore began by determining the sequence of this DNA (Fig. 1). We then identified the transcription start sites by performing nuclease S1 protection experiments using RNA from F9 cells differentiated to parietal endoderm-like cells (F9-PE) with cAMP and retinoic acid. We observed two major bands, of 151 and 73 nucleotides, which correspond to initiation 285 and 207 nucleotides upstream of the ATG, respectively (Fig. 2A). Similar results were obtained with RNA from F9 EC stem cells (data not shown). Analysis by primer extension experiments confirmed the nuclease protection data. We obtained two bands (Fig. 2B), the sizes of which, 179 and 101 nucleotides, again correspond to initiation 285 and 207 bases, respectively, upstream of the ATG. We therefore confirmed that there are two major transcription start sites in the Hoxb-4 promoter. The more distal of these is designated P1, and it is taken to define +1 in the sequence numbering (and as a consequence, the first nucleotide of the region under study is -926); the other promoter, P2, is thus located at +79 (Fig. 1). We also observed, in both the S1 and primer extension experiments, two weaker bands which map downstream of P2 and probably correspond to minor start sites. Additional primer extension experiments, using primers located close to the ATG, provided no evidence for other start sites. The sequences surrounding P1 and P2 contain no obvious TATA box or any GC-rich sequences. Database searches revealed that the 5' untranslated region had extensive homology with the chicken Hoxb-4 gene and with the mouse paralog Hoxa-4 (see Fig. 1A). In contrast, homology with the paralog Hoxa-4 was much more limited, and there was none to the paralog Hoxa-4.

Hoxb-4 5' flanking DNA contains two independent promoters. In order to identify the sequences required for Hoxb-4 transcription, we cloned the DNA from -926 to +285 upstream of a CAT reporter gene (Fig. 3) and transfected this construct into various cultured cell lines. CAT activity was determined and normalized with respect to that of pBLCAT2, a minimal thymidine kinase (TK) promoter construct, which was taken as 100% in each cell type. As an additional reference point the activity of a heat shock promoter construct (HSP-CAT) was measured. The -926/+285 construct was expressed in all cell lines tested (Fig. 4A), showing that this region has promoter activity. When cloned in the reverse orientation (+285/-926), this DNA was inactive (data not shown). The level of expression of the -926/+285 construct varied markedly in the different cell lines tested. Figure 4A shows that expression was low, 5% of TK, in HeLa cells; intermediate, 12

![FIG. 2. Identification of the transcription start site. (A) Nuclease S1 experiments. An EaeI restriction fragment containing the mouse Hoxb-4 promoter (+217 to +151) and plasmid sequences was end labelled with [γ-32P]ATP, and the antisense strand was isolated (see Materials and Methods) and used as a probe in S1 nuclease mapping experiments. A total of 150 μg of tRNA or total RNA from F9 cells differentiated with cAMP and retinoic acid was hybridized to this probe, and the products were treated with nuclease S1, fractionated in a 7 M urea-6% (wt/vol) polyacrylamide gel, and autoradiographed, as described in Materials and Methods. pBluescript KS+/HpaII radiolabelled fragments and products of sequencing reactions were used as markers. A diagram of the probe is shown: box, Hoxb-4 sequences; plain line, plasmid sequences. Major transcription start sites P1 and P2 are indicated by arrows, and minor transcription start sites are shown by shorter arrows. Sizes of probe and protected fragments as well as restriction enzyme sites employed are shown. (B) Primer extension experiments. A 65-mer oligonucleotide primer (+114/+179) was labelled with [γ-32P]ATP, hybridized to 100 μg of total RNA from F9 cells differentiated with cAMP and retinoic acid, and extended with reverse transcriptase, as described in Materials and Methods. The reaction products were fractionated in a 7 M urea-6% (wt/vol) polyacrylamide gel and autoradiographed. pBluescript KS+/HpaII radiolabelled fragments and products of sequencing reactions were used as markers.](https://example.com/fig2.png)
to 15% of TK, in fibroblasts (NIH 3T3 and 10T1/2), F9 EC cells, and differentiated PC12 (PC12D) cells; and high, 30 to 45% of TK, in F9-PE cells and the neuroblastoma cell lines SK-N-SH (SK) and Neuro2A (NEU). It is thus apparent that the Hoxb-4 promoter is subject to cell-type-specific regulation.

We next asked whether the two start sites, PI and P2, reflect the activities of two independent promoters. To do this, we generated constructs containing (i) both PI and P2 but lacking a negative regulatory element (−186/+285, see below), (ii) only PI (−186/+21), and (iii) only P2 (−21/+285). As expected, the constructs −926/+285 and −186/+285, which carry PI and P2, were expressed in F9 EC and Neuro2A cells (Fig. 4B). Interestingly, the constructs carrying only PI (−186/+21) or only P2 (−21/+285) were also efficiently expressed in both cell lines, showing that the promoter region of Hoxb-4 contains two separable promoters that can act independently.

A cell-type-specific, negative regulatory element is present between −226 and −186. Having defined the basic architecture of the Hoxb-4 promoter, we searched for regulatory elements, initially by constructing a series of 5′ deletions based on the −926/+285 construct (Fig. 3). Figure 5A shows that sequences between −926 and −226 had little, if any, effect on expression. However, deletion of the sequences between −226 and −186 led to a sixfold increase in expression in F9 EC cells, indicating the presence of a negative regulatory element (named element a). The effect of deleting this element is less marked in F9 EC-PE (3-fold), SK (2.6-fold), and Neuro2A (1.8-fold) cells, showing that it is regulated in a cell-type-specific fashion. Further deletions, up to −21, had little effect.

The 5′ untranslated region contains an element essential for promoter activity and a second, cell-type-specific, negative regulatory element. In order to explore the role of the 5′ untranslated region in the regulation of the Hoxb-4 promoter, we constructed three series of recombinants carrying 3′ deletions in this region. The −926 series contains all of the 5′ flanking DNA required for correct expression in the embryo; the −186 series lacks the negative regulatory element a defined above but includes both promoters, PI and PII; and the −21 series lacks both element a and PI but contains PII (Fig. 3). Analysis of the −926 series in F9 EC cells showed that sequences between +205 and +285 did not significantly affect expression (Fig. 5B). However, deletion of sequences between +169 and +205 led to a fivefold increase in CAT activity, thus defining a second regulatory element, element d. Deletion of the 20-bp fragment between +149 and +169 had a marked effect, reducing expression 10-fold, to very low levels, and thus defining an element, element c, that is essential for efficient promoter activity.

The −186 series of constructs exhibited essentially the same pattern of expression as the −926 series, although in each case the level of expression was higher because of the absence of the negative regulatory element a. These data show that elements c and d do not depend on element a for their function. The overall pattern of expression of the −21 series was again

![FIG. 3. Reporter recombinants. In the first line, a schematic diagram of the mouse Hoxb-4 gene promoter and 5′ noncoding region is shown (dashed lines, placid sequences; striped box, promoter sequences; white box, 5′ noncoding region). Numbering is from the first transcription start site (P1). Several restriction enzyme sites employed in the construction of reporter recombinants are also shown. In the second line, a model of the Hoxb-4 promoter and its regulatory elements a, b, c, and d is included. Stripped boxes, negative regulatory elements; grey box, positive regulatory element; thick boxes, cell-type-specific elements; arrows, P1 and P2 transcription start sites; P1 and PII, promoter I and promoter II, respectively; thin black boxes, transcriptional initiators. Below the second line, the schematic structures of the reporter recombinants are depicted. Arrows, transcription start sites; white boxes, −926 series of recombinants; black boxes, −186 series of recombinants; boxes with thin stripes, +21 series of recombinants; dark grey boxes, other Hoxb-4 recombinants; boxes with broad stripes, non-Hoxb-4 recombinants; light grey boxes, CAT gene; X, point mutations.](image-url)
recombinant was cotransfected with 4 µg of the internal control present between -226 and -186. A total of 12 µg of each reporter plasmid RSV-(βgal, by the calcium-phosphate method, into F9 EC
cells, F9 cells differentiated with retinoic acid and cAMP (F9-PE), and
the neuroblastoma cell lines SK-N-SH (SK) and Neuro2A. Cell
elements were transfected in F9 EC cells (B) and described in the text), were transfected in F9 EC cells (B) and 
Three series of recombinants, -926, -186, and +21 (depicted in Fig.
were prepared, and CAT activity was measured and corrected
by subtraction of background levels of the promoterless vector
extracts were prepared, and CAT activity was measured and corrected
by subtraction of background levels of the promoterless vector
pBLCAT3. Relative CAT activity was normalized with respect to that
of -926/+285, taken as 1.

FIG. 5. (A) A cell-type-specific, negative cis-regulatory element is
present between -226 and -186. A total of 12 µg of each reporter recombinant was cotransfected with 4 µg of the internal control plasmid RSV-βgal, by the calcium-phosphate method, into F9 EC
cells, F9 cells differentiated with retinoic acid and cAMP (F9-PE), and
the neuroblastoma cell lines SK-N-SH (SK) and Neuro2A. Cell extracts were prepared, and CAT activity was measured and corrected
by subtraction of background levels of the promoterless vector
pBLCAT3. Relative CAT activity was normalized with respect to that
of -926/+285, taken as 1. (B and C) Identification of negative and positive cis-regulatory elements in the 5' noncoding region of Hoxb-4.
Three series of recombinants, -926, -186, and +21 (depicted in Fig.
and described in the text), were transfected in F9 EC cells (B) and
and the neuroblastoma cell line Neuro2A (C). A total of 12 µg of each reporter recombinant was cotransfected with 4 µg of the internal control plasmid RSV-βgal, by the calcium-phosphate method. Cell extracts were prepared, and CAT activity was measured and corrected
by subtraction of background levels of the promoterless vector
pBLCAT3. Relative CAT activity was normalized with respect to that
of -926/+285, taken as 1.

similar, showing that elements c and d can regulate the P1l promoter in the absence of the PI promoter. We noticed the presence of an E box (CAGCTG), a consensus binding site for the proteins of the basic helix-loop-helix family of transcription factors, at nucleotides +166 to +172. However, mutation of this binding site had no effect on expression (compare +21/+285 mut with +21/+285 in Fig. 5B).

When the same series of recombinants were analyzed by using Neuro2A cells (Fig. 5C), we found that deletion of region d had no effect, indicating that its activity, like that of the other negatively acting element, element a, is regulated in a cell-type-specific fashion. However, the positively acting element c was, as in F9 EC cells, essential for efficient expression.

Both PI and PII contain transcriptional initiators, the activity of which is modified by neighboring sequences. In order to identify the minimal sequences required for transcriptional activity, additional mutants were derived from the virtually inactive reporters -186/+149 and +21/+49, and their activities were assayed by using Neuro2A cells (Fig. 6A). Removal of sequences between +113 and +149 had no effect, but deletion of the region between +21 and +113 restored expression, and consequently, a third negative element, element b, is delimited. As deletion of sequences from +21 to +46 and those from +86 to +113 rendered PII active again (see construct +46/+86 [Fig. 6A]), element b must be located in one of these two regions. Smaller constructs encompassing only the PI promoter showed lower levels of expression (see -32/+21 and -5/+7), but even a 12-bp element (+5/+7) was active, suggesting that it acts as a transcriptional initiator (35, 36). Similarly, the -75/+86 construct encompassing 12 bp of PI1 gave a lower level of expression, suggesting that PII also includes an initiator. The region surrounding these two initiators contains potential binding sites for the transcription factor Spl. However, mutation of these sites had no effect on expression (see -32/+21 mut and +46/+86 mut).

Elements b, c, and d in the Hox-b 4 5’ untranslated region regulate mRNA levels. The location of elements b, c, and d in the 5’ untranslated region raises the possibility that they regulate expression at the level of translation rather than transcription. We therefore performed quantitative nuclear
S1 experiments using RNA extracted from Neuro2A cells (for elements b and d) or from F9 EC cells (for element d) that had been transfected with the appropriate reporter (see Fig. 5B).

The promoterless vector pBLCAT3 gave rise to no detectable
mRNA, while the data with the other constructs paralleled those obtained from the CAT assays. In F9 EC cells, deletion of region d increased the amount of transcript sixfold. In Neuro2A cells, deletion of region c reduced the level of transcript 15-fold, while deletion of region b increased it 6-fold. These data show that the three elements in the 5’ untranslated region act by regulating mRNA levels.

Regulatory element c activates transcription of heterologous promoters. To further characterize positive regulatory element c, we cloned it into CAT-based reporters, upstream of four different promoters, in both orientations. Two of these promoters are heterologous. One is essentially a TATA box (19) which drives a very low level of transcription (the TATA box series [Fig. 7A]); the other is the TK minimal promoter (20) with a much higher level of basal transcription (the TK promoter series). The other two promoters belong to the Hox-b 4 gene and are characterized in this report. One is initiator 1, which contains the -5 to +7 sequences of the gene (Fig. 1 and 6A) and which drives a very low level of transcription, comparable to that of the TATA box-based recombinants (the Hox-b 4 initiator 1 series). The other is the minimal promoter that comprises the -32 to +21 sequences (Fig. 1 and
Voxl were transfected into Neuro2A cells. Transfection and CAT assays with those without it, irrespective of the orientation of the total RNA was hybridized with a probe specific for CAT mRNA pBCTBX2 (15). A total of 150 μg (Neuro2A) or 200 μg (F9 EC) of different kinds of recombinants were used: the -186, the +21, and the minimal promoter alone (-32/+21) and 2.5 times higher than the longer version of element c, suggesting that the +135 to +143 sequences have a negative effect on transcription. These experiments further delimited element c (+143 to +169) and show that it activates transcription in an orientation- and promoter context-independent manner.

Two different factors bind independently to overlapping binding sites in element c. To characterize the factor(s) that activates through element c, we performed DNA mobility shift experiments with oligonucleotides comprising the +135 to +169 or the +143 to +169 sequences (see Fig. 7B). With the smaller probe (ACT-WT) we detected two major bands, A and B (lanes 2 and 5), that are specifically competed for by an excess of cold oligonucleotide (lanes 3 and 4 and 6 and 7). With the larger probe (LACT), band A is partly replaced by band X, which migrates slightly faster (lane 1) and which was not further characterized. Factors A and B have a wide distribution. Both are present in Neuro2A (lanes 2 to 4) and F9 EC (lanes 5 to 7) cells. Interestingly, they are also present in 10.5-day mouse embryos in a variety of different structures: forebrain (lane 8), hindbrain (lane 9), branchial arches (lane 10), liver (lane 11), and heart (lane 12), as well as in 8.5- and 14-day embryos (data not shown).

To determine the exact binding sites of factors A and B, we systematically mutated the +135 to +169 sequences and performed mobility shifts with the mutant oligonucleotides (Fig. 8). These experiments revealed two overlapping binding sites with similar but not identical specificities. While M3 is unable to bind any of the factors, M5 specifically binds factor A but not factor B. In contrast, M8 binds factor A with greatly diminished affinity, but the binding of factor B remains unchanged (or is even enhanced). These results, and the existence of similar binding sites in the myogenin and myoD1 promoters, allowed us to establish a consensus for each factor (Fig. 8); G C C A T/G T G/C G/C for factor A and C A C A T T T for factor B.

We also competed for binding to the wild-type probe with sites that showed some sequence similarity to those defined above. While NF-1 and NFy sites did not compete (data not shown), a YY1 binding site did, but only for the binding of factor B (Fig. 9B, compare lanes 1 and 4) strongly suggesting that factor B is indeed YY1 and is unrelated to factor A. In contrast, the myogenin binding site competed for the binding of only factor A (compare lanes 1 and 3), showing again that the two factors bind independently to the Hoxb-4 sequences. In conclusion, element c binds two factors: one is likely to be the YY1 transcription factor; the other does not seem to be related to any known DNA-binding protein, and we will call it HoxTF (for homeobox gene transcription factor).

The HoxTF binding site mediates activation of transcription of the Hoxb-4 minimal promoter. To distinguish which of the two factors, YY1 or HoxTF, is the activator, we introduced mutations specific for each of them and for both in combination. The double mutation completely abolished activation by element c (Fig. 9A, compare Double mut with ACT-WT). Mutation of the HoxTF site had a similar although less drastic effect (compare HoxTF mut with ACT-WT). The residual activity of HoxTF mut is most likely explained by the residual binding of the factor by the mutant sequence (10 to 20% of wild type; see M8 in Fig. 8), as can be noticed in Fig. 9B, which shows that HoxTF mut, but not Double mut, partially inhibits the HoxTF band. In contrast to HoxTF mut, mutation of the YY1 site had only a slight effect. These experiments show that the HoxTF binding site activates transcription of the Hoxb-4 minimal promoter.
The Hoxb-4 gene has two independent promoters active in a variety of cell types. We have shown that the sequences immediately upstream of the Hoxb-4 gene have promoter activity in that they are able to drive transcription of the CAT reporter gene in a variety of cultured cell lines. There are considerable differences in the activity of the promoter in the cell lines tested, and it is of interest that these differences reflect, in general terms, the transcriptional activity of the endogenous gene extrapolated from data from both the embryo and cell lines. Hoxb-4 expression is high in neuronal cells (neuroblastoma cell lines Neuro2A and SK-N-SH) and parietal endoderm cells (F9-PE) and low in epithelial cells (HeLa), with intermediate levels in fibroblasts (NIH 3T3 and 10T1/2) and neurons derived from the neural crest (differentiated PC12).

The 5' flanking sequences contain two transcription start sites, P1 and P2, 285 and 207 bases upstream of the translation initiation codon, respectively. Constructs containing either P1 or P2 alone are efficiently expressed, showing that the Hoxb-4 gene has two independent promoters (P1 and P2). In each case we have shown that a 12-bp element, centered on the start site, retains activity, suggesting that both promoters act as transcriptional initiators (35, 36). This idea is consistent with the absence of TATA boxes or GC-rich sequences.

Negative regulation of the Hoxb-4 promoter. Our mutational analysis has identified three elements that diminish the activity of the Hoxb-4 promoter (Fig. 10B). Element a, located between −226 and −186, acts negatively in a cell-type-specific fashion. Deleting it increases expression sixfold in F9 EC cells but has less effect in F9-PE and neuroblastoma cells. The other two elements are located in the 5' untranslated region, but they clearly affect mRNA levels. Element d (+169 to +205) is similar in its properties to element a. It acts negatively in F9 EC cells but is without effect in neuroblastoma cells. In contrast, element b does negatively regulate expression in neuroblastoma cells. The negative elements a and d act additively and both can regulate each of the promoters. They are presumably responsible for the inefficient expression of the Hoxb-4 promoter in F9 EC cells compared with F9-PE and neuroblastoma cells. Arcioni et al. have shown that the HOXC-5 promoter contains a negatively acting element, of similar cell specificity, located upstream of the transcription start site (1). Moreover, Puschel et al. have demonstrated that the Hoxa-7 gene contains elements within the 5' untranslated region that restrict the activity of the promoter in the embryo (26). The presence of negative regulatory elements could explain the activation of Hox gene expression in EC cells by inhibitors of protein synthesis (32).

The Hoxb-4 gene has two independent promoters active in a variety of cell types. We have shown that the sequences immediately upstream of the Hoxb-4 gene have promoter activity in that they are able to drive transcription of the CAT reporter gene in a variety of cultured cell lines. There are considerable differences in the activity of the promoter in the cell lines tested, and it is of interest that these differences reflect, in general terms, the transcriptional activity of the endogenous gene extrapolated from data from both the embryo and cell lines. Hoxb-4 expression is high in neuronal cells (neuroblastoma cell lines Neuro2A and SK-N-SH) and parietal endoderm cells (F9-PE) and low in epithelial cells (HeLa), with intermediate levels in fibroblasts (NIH 3T3 and 10T1/2) and neurons derived from the neural crest (differentiated PC12).

The 5' flanking sequences contain two transcription start sites, P1 and P2, 285 and 207 bases upstream of the translation initiation codon, respectively. Constructs containing either P1 or P2 alone are efficiently expressed, showing that the Hoxb-4 gene has two independent promoters (P1 and P2). In each case we have shown that a 12-bp element, centered on the start site, retains activity, suggesting that both promoters act as transcriptional initiators (35, 36). This idea is consistent with the absence of TATA boxes or GC-rich sequences.

Negative regulation of the Hoxb-4 promoter. Our mutational analysis has identified three elements that diminish the activity of the Hoxb-4 promoter (Fig. 10B). Element a, located between −226 and −186, acts negatively in a cell-type-specific fashion. Deleting it increases expression sixfold in F9 EC cells but has less effect in F9-PE and neuroblastoma cells. The other two elements are located in the 5' untranslated region, but they clearly affect mRNA levels. Element d (+169 to +205) is similar in its properties to element a. It acts negatively in F9 EC cells but is without effect in neuroblastoma cells. In contrast, element b does negatively regulate expression in neuroblastoma cells. The negative elements a and d act additively and both can regulate each of the promoters. They are presumably responsible for the inefficient expression of the Hoxb-4 promoter in F9 EC cells compared with F9-PE and neuroblastoma cells. Arcioni et al. have shown that the HOXC-5 promoter contains a negatively acting element, of similar cell specificity, located upstream of the transcription start site (1). Moreover, Puschel et al. have demonstrated that the Hoxa-7 gene contains elements within the 5' untranslated region that restrict the activity of the promoter in the embryo (26). The presence of negative regulatory elements could explain the activation of Hox gene expression in EC cells by inhibitors of protein synthesis (32).
on cell type, although in its absence the promoter retains detectable activity (4% of TK). This element can activate transcription of its own promoter as well as that of heterologous promoters, in an orientation-independent manner and in different contexts. It is able to interact with single elements like a TATA box or an initiator, but it is also functional when placed upstream of more complex promoters that exhibit a TATA box or an initiator, but it is also functional when placed upstream of more complex promoters that exhibit several regulatory elements. Its versatility is also exemplified by its normal location is in the 5' untranslated region, downstream of the transcription start site.

A long version of element c (+135 to +169) is less effective than a shorter one (+143 to +169) in activating transcription, indicating that the +135 to +143 sequences have a negative effect. Two factors bind to overlapping sites within the +143 to +169 region. One seems to be related to the YY1 transcription factor, as shown by our mutational analysis and by comparison with similar binding sites in the myogenin and myoD1 genes and consensus binding sites for bands A and B are also presented. Binding activities of the mutant sequences, compared with that of the wild type, are indicated for both complexes A and B. Sequences of similar binding sites in the mouse myogenin and myoD1 genes and consensus binding sites for bands A and B are also presented.

FIG. 8. Binding activity of element c. End-labelled, double-stranded oligonucleotides containing wild-type (WT) or mutated sequences (M1 to M10) of Hoxb-4 (+143 to +169) were used as probes in gel shift experiments as described in Materials and Methods. Mutated bases are circled. Whole-cell extracts prepared from the neuroblastoma cell line Neuro2A were used as a source of proteins. Binding activities of the mutant sequences, compared with that of the wild type, are indicated for both complexes A and B. Sequences of similar binding sites in the mouse myogenin and myoD1 genes and consensus binding sites for bands A and B are also presented.

FIG. 9. Mutation of the HoxTF binding site abolishes activation of transcription by element c. (A) Transcription. Several recombinants were transfected into the Neuro2A cell line: the −32/+21 and ACT-WT reporters were described in the legend to Fig. 7. In the HoxTF mut, YY1 mut, and Double mut reporters, the mutated sequences M8, M5, and M3 (Fig. 8) replaced the wild-type sequences of ACT-WT. A total of 16 μg of each reporter recombinant was cotransfected with 4 μg of the internal control plasmid RSV-(3 gal, by the calcium-phosphate method. Cell extracts were prepared, and CAT activity was measured and corrected by subtraction of background levels of the promoterless vector pBLCAT3. Relative CAT activity was normalized with respect to that of −32/+21, taken as 1. MPB4, Hoxb-4 minimal promoter. Broken arrows indicate the start site of transcription. (B) Binding activity. An end-labelled, double-stranded oligonucleotide containing wild-type sequences of Hoxb-4 (+143 to +169) was used as a probe in gel shift experiments as described in Materials and Methods. Whole-cell extracts prepared from the neuroblastoma cell line Neuro2A were used as a source of proteins. Specific competitor, as indicated, was used at 100-fold molar excess. Act-wt, YY1 mut, HoxTF mut, and Double mut are wild-type and mutated sequences (+143 to +169) from the Hoxb-4 promoter (WT, M5, M8, and M3, respectively, in Fig. 7). Myo comprises sequences from +9 to +35 of the mouse myogenin promoter (40); the YY1 binding site is described by Ellis et al. (10). HoxTF and YY1 indicate the two main retarded complexes (the same as bands A and B in Fig. 7 and 8). FREE, unbound probe.
effect. In conclusion, HoxTF is the active factor for the positive regulatory element c and is essential for efficient expression of the Hoxb-4 gene. The element does not show any cellular specificity, which is in agreement with the wide distribution of the factor, in tissue culture cell lines as well as in the mouse embryo.

Our transgenic experiments showed that the 1.2-kb region of Hoxb-4 under study here has no activity in the mid-gestation embryo, with the exception of ectopic expression in the superior colliculi (38). However, it is activated by spatially specific enhancers located both within the gene and downstream of it. In contrast, as reported here, this region is active in a variety of cell lines. Perhaps these cell lines reflect, in part, the situation in the later embryo or in the adult. A factor(s) that binds within the 1.2-kb region must be able to activate the gene. Among the regulatory elements we have characterized, the HoxTF binding site, the only one that is essential for efficient transcription of the Hoxb-4 gene, is the most likely candidate. Interestingly, analogous switches seem to exist in some D. melanogaster homeobox genes (see next section).

Do homeobox gene promoters have common features? The main structural and functional features of the Hoxb-4 promoter are diagrammed in Fig. 10B and can be summarized as follows. (i) The minimal promoter contains two initiators which act independently to dictate specific, albeit inefficient, initiation of transcription. (ii) The upstream sequences contain a cell-type-specific, negative regulatory element (element a) and sequences around each initiator which increase efficiency. (iii) The downstream sequences contain two negative regulatory elements (elements b and d), the latter of which is cell type

FIG. 10. (A) Sequence comparisons between the mouse Hoxb-4 5′ noncoding region and homologous and paralogous genes. Sequence alignment between 5′ noncoding regions of mouse Hoxb-4 (MM Hox-B4), chicken Hoxb-4 (CH Hox-B4 [30]), mouse Hoxc-4 (MM Hox-C4 [13]), chicken Hoxd-4 (CH Hox-D4 [31]), and mouse Hoxd-4 (MM Hox-D4 [25]). The conserved regulatory elements (see Results) are indicated: black bar, HoxTF binding site; grey bar, element d. (B) Model of the mouse Hoxb-4 gene promoter. Schematic diagram of the mouse Hoxb-4 gene promoter and its regulatory elements (a, b, c, and d). Location of element b is uncertain; it could be between +21 and +47 or (as shown) between +85 and +113. Numbering is from the PI transcription start site. The HoxTF binding site is shown in its approximate location. Asterisks indicate essential for binding.
specific, and a powerful positive regulatory element (element c) which acts in all cell types tested.

The promoters of three D. melanogaster homeobox genes, Ultrabithorax, Engrailed, and Antennapedia (P2), lack TATA boxes andSpl-like binding sites and appear to depend on initiators. All three have positive regulatory elements in the 5' nontranslated region that are essential for transcription in vitro in embryonic extracts (4, 24, 37). These general similarities to the *Hox*-4 promoter are not reflected in any obvious sequence homologies.

The promoters of mammalian homeobox genes are not well characterized. Nonetheless, it appears that there are several different types of minimal promoter. Some genes, e.g., *HOXC*-5 (1) and *H OXD*-4 (8), have putative TATA boxes, while others, e.g., *Hoxa*-4 (11) and *Hoxc*-4 (13), that lack TATA boxes contain, between -60 and -30, potential Spl binding sites that are known to be able to mediate the initiation of transcription at multiple, closely spaced sites. A third class of genes, e.g., *Hoxc*-6 (9) and *Hoxc*-8 (2), lack such sequence motifs and thus may, like Hoxb-4, depend on initiators. It is striking that promoter architecture appears to be quite different even among the genes of a paralogous set. However, it is noteworthy that the negative element d is conserved in the paralogs *Hoxc*-4 and Hoxd-4, although not in Hoxa-4, suggesting that there are regulatory mechanisms common to some genes in the paralogous group.

Homeobox gene promoters do not appear to share a common organization. Nevertheless, it is quite clear that a large subset of them share some unusual features. They are often multiple, sometimes close together, sometimes spaced up to several kilobases apart (7, 8, 33). A relatively high proportion contain minimal promoters that do not have TATA boxes but have initiators instead. They can be regulated by elements situated upstream or, interestingly, downstream of the transcription start site. The *Hox*-4 promoter has all three of these characteristics. Its complex and unusual organization is appropriate for a developmental control gene, exhibiting precise spatial and temporal expression patterns, and is thus subject to sophisticated transcriptional regulation.

**Conclusion.** The experiments reported here have enabled us to define the structural and functional architecture of the *Hox*-4 promoter and give insights into how other developmentally regulated genes may be organized. We can now investigate, with transgenic animals, how the numerous regulatory elements and the factors that bind to them interact with the spatially specific enhancers of the gene, which must set the precise boundaries of expression along the anteroposterior axis in both the mesoderm and the hindbrain, and can ensure that the levels of gene product made are appropriate to its role in the patterning process.

**Acknowledgments**

We thank Robb Krumlauf for the gift of the Hox-4 clone and for helpful comments on the manuscript. We also thank Rudi Allemann for advice and the members of the Rigby laboratory for useful discussions.

J.G. holds a Research Studentship from the Medical Research Council, which also paid for this work.

**References**


element of the murine Hox-4.2 gene. EMBO J. 11:3673-3680.
  tory Press, Cold Spring Harbor, N.Y.
  74:5463-5467.
31. Sasaki, H., E. Yokoyama, and A. Kuroiwa. 1990. Specific DNA binding of the two chicken Deformed family homeodomain pro­
Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*

**SAMUEL APARICIO**, **ALASTAIR MORRISON**, **ALEX GOULD**, **JONATHAN GILTHORPE**, **CHITRITA CHAUDHURI**, **PETER RIGBY**, **ROBB KRUMLAUF**, AND SYDNEY BRENNER

Molecular Genetics Unit, Department of Medicine, Addenbrookes Hospital, Cambridge CB2 0QO, United Kingdom; Laboratory of Developmental Neurobiology, Laboratory of Eukaryotic Molecular Genetics at National Institute for Medical Research. The Ridgeway Mill Hill, London NW7 1AA, United Kingdom

**ABSTRACT** Comparative vertebrate genome sequencing offers a powerful method for detecting conserved regulatory sequences. We propose that the compact genome of the teleost *Fugu rubripes* is well suited for this purpose. The evolutionary distance of teleosts from other vertebrates offers the maximum stringency for such evolutionary comparisons. To illustrate the comparative genome approach for *F. rubripes*, we use sequence comparisons between mouse and *Fugu Hoxb-4* non-coding regions to identify conserved sequence blocks. We have used two approaches to test the function of these conserved blocks. In the first, homologous sequences were deleted from a mouse enhancer, resulting in a tissue-specific loss of activity when assayed in transgenic mice. In the second approach, *Fugu* DNA sequences showing homology to mouse sequences were tested for enhancer activity in transgenic mice. This strategy identified a neural element that mediates a subset of *Hoxb-4* expression that is conserved between mammals and teleosts. The comparison of noncoding vertebrate sequences with those of *Fugu*, coupled to a transgenic bioassay, represents a general approach suitable for many genome projects.

The major challenges for all genome projects will be deciphering gene function and regulation. In the absence of useful grammars, testing the value of sequences by transgenesis is an important approach to finding regulatory information buried in the noncoding regions that constitute 95% of mammalian genomes; however, transgenic analysis is costly and labor intensive, and means of facilitating this approach are required. We have shown (1) that the Japanese puffer fish *Fugu rubripes* (*Fugu*) has a compact vertebrate genome and propose that one use of *Fugu* is the discovery of conserved gene regulatory sequences by comparison of *Fugu* noncoding sequences with those of higher vertebrates. The *HOM/Hox* complexes are some of the best studied cases of structural, functional, and regulatory conservation (2-5), and we therefore chose *Fugu Hox* genes as examples to study the degree of conservation in gene regulation. As part of a systematic analysis of the *Fugu Hox* gene clusters, the *Hoxb-4* gene was cloned and sequenced, since the regulation and function of this gene have been examined in detail in mice (6-8). Here we show that sequence comparisons between *Fugu* and mouse can be used to identify conserved regulatory elements that direct subsets of the full *Hoxb-4* expression pattern in transgenic mice.

**MATERIALS AND METHODS**

Cloning and Sequencing of *Fugu Hoxb-4*. Degenerate PCR primers hox26l1 and hox26f1 whose sequences were derived from the human and mouse *Hoxb-4* exon 2 sequences were used to amplify and clone a fragment of *Fugu Hoxb-4* from *Fugu* genomic DNA. *Taq* DNA polymerase (Cetus) was used with a Mg\(^2+\) concentration of 1.5 mM final. A *A2001 Sau3AI Fugu* genomic library (1) was screened at high stringency with this probe by using the method of Church and Gilbert (9). A phage designated H26S6A1 was obtained after three rounds of plaque purification that contained the *Hoxb-4* gene. The region from an internal *EcoRI* restriction site to the λ polylinker, approximately 11.4 kb, was subcloned and sequenced by the dideoxynucleotide method (10) on an ABI373A DNA sequencer according to the manufacturer's instructions.

**DNA Sequence Analysis.** The conceptual translation of *Fugu* and mouse *Hoxb-4* DNA sequences (see Fig. 1) was aligned by using CLUSTALV (11). Regions of nucleotide sequence similarity were detected with two methods. First, pairwise DNA alignments were made by using X-SNP in the Staden suite of programs (12). The scoring matrix used was the default MDM78, the identity score was 6, the odd-span length was 9, the proportion score was taken as default, and the standard deviation above the mean was 2.5. Regions of similarity appear as continuous lines on the plots. These subregions were then either locally aligned by using the Smith-Waterman algorithm or searched for the best local similarity (13) by using the programs high, medium, and low (J. Parsons, personal communication) to search larger tracts of sequence for short similarity blocks.

**Generation of Transgenic Mice.** Regulatory regions were inserted into two basic lacZ reporter constructs that used either the *hsp68* promoter (610Za; ref. 6) or the *Hoxb-4* promoter (construct 8; ref. 6). Construct 1 contains the mouse 1.38-kb Sal I-Bgl II fragment of *Hoxb-4* (wild type). Construct 2 contains a 1.28-kb Sal I-Bgl II fragment from mouse *Hoxb-4*, where the conserved region (CR1) was deleted by removing the internal Mun I-Sfi I fragment. Construct 3 contains a 3-kb HindIII-Neo I mouse *Hoxb-4* fragment. Construct 4 contains a 5.1-kb *Kho I-Sal I* fragment from phage H26S6A1 containing *Fugu Hoxb-4*. Construct 5 contains the 250-bp conserved CR2 region generated by PCR with oligonucleotides (BAM- TCTTGGTTATGGTCTTAAAAATGC and BAM-CTCTGCTCTATATTATTTTCCGATAC; BAM, a BamHI restriction enzyme recognition sequence for cloning) from phage H26S6A1. Construct 6 contains the 300-bp CR3 region generated by PCR with oligonucleotides (BAM-TCTTATATAATGAATTGGGAGCAG and BAM- GGGTGATGGGAGGCTG) from phage H26S6A1. All inserts were released from vector sequences prior to microinjection.

Animal experiments were performed according to procedures regulated under the project licences issued by the Home Office and the Genetic Manipulation Advisory Group. The generation and analysis of transgenic mice from F1 hybrids (CBA × C57BL/6) were as described (6). Pseudopregnant females were injected intraperitoneally with the anaesthetic Avertin at a dose of 0.015-0.017 ml of a 2.5% solution per gram.

Abbreviations: *rn*, rhombomere number; *dpc*, days post coitum.

1To whom correspondence and reprint requests should be addressed.
of body weight prior to embryo transfer. Embryos were photographed with the aid of a Leitz M10 stereomicroscope and camera. The numbers of embryos in each experiment are expressed in the figure legends as follows: the number of embryos with the pattern shown/total number of embryos expressing lacZ.

RESULTS

Sequence Comparisons. Conceptual translation and comparison of the cloned Fugu DNA sequence with the mouse Hoxb-4 sequence reveals multiple regions of similarity over the entire protein, confirming the Fugu sequence as Hoxb-4 (Fig. 1A). The pairwise identity score was 56% over the whole protein but approached 100% in the homeodomain region and 90% in the N-terminal region diagnostic for paralogs in group 4. The putative splice junction corresponds in phase and position to the mouse sequence. At the nucleotide level the sequence was closer to b-4 than to other group 4 paralogs (data not shown).

Most Fugu genes are smaller than their mammalian counterparts because of expanded introns (ref. 1 and unpublished observations). However, in Hoxb-4 the introns in Fugu and mammals are of a similar size; we presume that this is because of functional constraints during evolution due to the presence of transcriptional regulatory elements within the intron. Sequence comparisons between the intron and 3' noncoding regions of mouse and Fugu Hoxb-4 revealed three positionally conserved islands of sequence homology (Fig. 1B and C). It is intriguing that these sequences mapped within two large noncoding regions of mouse Hoxb-4 previously shown to contain enhancers that direct major aspects of the spatially restricted patterns of expression (6). The first conserved island, CR1, mapped within the region-C enhancer. Region C mediates the proper spatially restricted expression pattern of Hoxb-4 in the mesoderm and gives a subset of the full Hoxb-4 pattern in the central and peripheral nervous systems. The second and third conserved islands, CR2 and CR3, are located within the region-A enhancer. This 3-kb region, 3' of mouse Hoxb-4, directs expression in the neural tube with a sharp anterior boundary at the junction of rhombomeres 6 and 7 (r6/7) in the hindbrain (Fig. 3A and B and ref. 6). Posterior to r6/7, expression extends back along the entire length of the spinal cord.

Testing of Conserved Sequences. We have used two approaches to assess whether these conserved islands of sequence homology are functionally significant. In the first approach, we examined whether the conserved Fugu CR1 sequence was able to direct us specifically to critical elements within a larger mouse enhancer. The 92-bp conserved CR1 region was internally deleted from the mouse region-C enhancer (Fig. 2A), and the remaining sequence was functionally tested in a lacZ reporter gene construct in transgenic mice. The major domains of expression in mesoderm and ectoderm observed with the wild type (Fig. 2B; construct 1) were lost; however, there was a consistent domain of residual expression in the ventral neural tube, showing that the construct could still function at multiple integration sites (Fig. 2C; construct 2). This demonstrates that CR1 is an essential component for the correct expression in the mesoderm and the central and peripheral nervous systems. Thus, sequence homology has identified at least one required regulatory component within the region-C enhancer.

In a second approach we tested the ability of Fugu sequences to direct expression in transgenic mice. Both of the homology regions CR2 and CR3 lie within a region 3' to Fugu Hoxb-4 that is topologically equivalent to the mouse neural enhancer (region A; construct 3). We first tested the ability of a 5-kb fragment 3' of Fugu Hoxb-4 (construct 4) to direct expression and observed staining in the neural tube. Furthermore, the anterior boundary of expression was precisely at the r6/7 junction in the hindbrain, the same anterior boundary seen with the mouse region-A construct (Fig. 3). Using stable transgenic mouse lines, we also observed that the temporal progression of neural expression mirrored that observed with the mouse regulatory region (data not shown). There were differences between mouse and Fugu transgenic constructs. With the Fugu transgene we detected the presence of posterior mesoderm expression but no expression in the neural tube below the midhindbrain region. The region of Fugu we tested was considerably larger than mouse region A, and it is possible that additional elements relating to Hoxb-3, which could account for the mesodermal expression, had inadvertently been included. For example, the mouse region A also contains a
Fig. 2. Homology defines a critical regulatory component of the region-C enhancer. (A) Mouse Hoxb-4 region-C DNA and the equivalent region from Fugu were aligned on the area of intron homology (stippled boxes), the sequences of which are shown below. Sequence identities are highlighted in solid black. Coding regions are represented by hatched boxes. (B) Pattern of expression from wild-type (construct 1) mouse region-C enhancer in a 10.5 days post coitum (dpc) transgenic mouse embryo (5/5 transgenes). (C) Pattern of expression in an 11.5-dpc transgenic mouse embryo where CR1 (92 bp) has been internally deleted from region C (construct 2). Note residual expression in ventral neural tube (6/6 transgenes).

Sa, Sal I site; Bg, Bgl II site; M, Mun I site; and S, Sfi I site.

promoter for the Hoxb-3 gene (14). In support of this, the anterior mesoderm expression is more anterior than that observed for Hoxb-4 (6, 7, 14).

We then examined whether CR2 and CR3 are involved in the conserved aspects of neural tube expression directed by the 3′ region of the Fugu Hoxb-4 DNA (Fig. 4). A 250-bp fragment containing only CR2 was tested in the transgenic assay but gave no consistent expression pattern in transgenic mice (construct 5; data not shown). However, the 300-bp fragment encompassing CR3 directs expression confined to the posterior hindbrain, primarily in r7 and r8 (Fig. 4; construct 6). The anterior expression limit at the boundary between r6 and r7 (Fig. 4 C and D) was identical to that seen with both the full-length mouse region A and the 5-kb Fugu construct (Fig. 3 C and D). We note that the posterior extent of expression was variable among the different constructs, which could reflect missing or evolutionarily divergent elements. Posterior expression limits also vary between species, which is not surprising given that it is the anterior expression domains that are the key determinants of the patterning effects of Hox genes (2–5, 15, 16). This experimental approach has demonstrated that Fugu sequences can function in transgenic animals and that such comparisons, when linked to a bioassay, are useful for defining gene regulatory motifs.

DISCUSSION

Teleosts are the most distant extant vertebrate precursors of mammals, with a separation time of 430 million years (17), and offer the maximum stringency for genome sequence comparisons among vertebrates. All unconstrained sequences have had the maximum time to randomize by mutation, and only those sequences required for functions common to all vertebrates will be conserved. This is true not only for coding sequences but also for sequences involved in the regulation of gene expression. That at least some aspects of Hox gene expression have been highly conserved is consistent with this view.

Evolution of genomes and of developmental mechanisms has proceeded in parallel. Comparisons of mammalian sequences with those of other vertebrates in an intermediate taxonomic position are valuable, as illustrated by mouse/avian sequence comparisons for the posteriorly expressed gene Hoxd-11 (18). (The avian/mammal divergence occurred 220 million years ago.) Moreover, they may offer insights into regulatory processes that arose later in evolution and are linked to a particular developmental strategy. However, intermediate divergence times will also provide less randomization of unconstrained sequences and potentially more homologous sequences whose function cannot be assigned.

Three-way comparisons are likely to prove informative for identifying sequences involved in universal aspects of regulation.
as well as in mechanisms shared only by organisms that develop similarly. Although the native expression directed by Fugu CR3 is unknown, the strong regulatory conservation observed, which is sufficient to direct the anterior expression domain of the hindbrain in the mouse, is consistent with conservation of anterior function as a universal property of Hox genes (19). There will be other cases of conservation and divergence within a developmental process. For example, it is known that during skeletal muscle development four basic helix-loop-helix factors play crucial roles in myogenesis. Their expression is restricted to skeletal muscle cells in all vertebrates, although they are deployed in different orders in mammals, birds, and frogs (20). Therefore, one would expect to find conservation of tissue specific but not temporal regulatory elements.

Transgenesis in other vertebrate species, such as teleosts like the zebra fish, may offer intermediate assays in evolutionary terms for testing higher vertebrate genes. For example, there is some evidence for conservation of the function of murine regulatory regions in zebra fish (21). We expect that comparisons at the sequence and functional levels between other teleosts, such as zebra fish, will prove informative by analogy to insights gained for chorion genes between moths and Drosophila (22, 23) and for ADH/GLD genes between Drosophila melanogaster and other Drosophila species (24, 25).

The uniquely compact properties of the Fugu genome aid in sequence comparisons, as similarities are not obscured by the large amounts of dispersed repetitive DNA found in other genomes. We show that noncoding sequence homologies clearly exist and have heuristic value as predictors of regulatory elements. It is interesting that regions as large as 100–200 bp are conserved in the enhancers, since such regions would be capable of binding multiple transcription factors. This suggests that the interaction of several factors required to regulate spatially restricted gene expression has been conserved throughout vertebrate evolution. Such regulatory conservation is unlikely to be specific to Hox genes, as many developmental processes appear conserved in evolution, and we expect that, at least in these cases, regulatory information will also be conserved. Where homology blocks are found in other gene classes, the task of analyzing the cognate mammalian regions will be simplified from testing deletions of kilobase regions down to segments of a few hundred base pairs or less.

The existence of very short, conserved sequences would be difficult to detect with present methods. However, when information exists about the nature of the transactivating factors, short consensus elements may be searched for between two or more species. This approach, used in parallel with a transgenic assay, has successfully revealed the existence of an essential and conserved DR2 type retinoic acid receptor binding site 3′ to the Hoxb-1 loci of Fugu, chick, and mouse (26). The systematic comparative sequencing approach demonstrated here is providing similar insights for other classes of genes that are also being investigated. We propose this use of Fugu as a general approach for all genome projects.

S.A. and A.M. were supported by a Medical Research Council Training Fellowship, C.C. was supported by a Human Frontiers Science Program collaborative grant, A.G. was a Beit Memorial Research Fellow, and J.G. was supported by a Medical Research Council Studentship.

Fig. 4. *Fugu* CR3 is sufficient to direct the anterior rhombomere-restricted expression of *Hoxb-4*. (A) Alignment of mouse (M) and *Fugu* (F) 3' regions. Exon 2 is shown as a hatched box with a solid center representing the homeodomain. The homology regions CR2 and CR3 are indicated as shaded boxes. The sequence of CR3 is shown beneath the *Fugu* locus, with conserved residues highlighted in uppercase, boldface type. Constructs 5 and 6 are marked as horizontal lines above the *Fugu* locus. Restriction sites: Bg, Bgl II; H, HindIII; N, Nco I; Sa, Sal I; and X, Xba I. (B) Dorsal view of 10.5-dpc transgenic mouse embryo carrying *Fugu* region A (construct 4). A sharp anterior boundary of expression is seen in the neural tube at the r6/7 junction. (C) Dorsal view of 10.5-dpc transgenic mouse embryo carrying *Fugu* CR3 (construct 6). Expression is seen primarily in r7. Expression is seen primarily in r7. Note the sharp boundary of expression at the r6/7 border (indicated). The most anterior staining in the superior colliculus is due to the basal promoter alone (6). (D) Flat mount of hindbrain of embryo in C. The anterior boundary of expression at r6/7 is indicated by the dashed line. There is no sharp posterior boundary of expression. Consistent patterns were obtained in 1/3 expressing embryos transgenic for CR3. The positions of the otic vesicles (ov) lying adjacent to r5 and r6 are indicated by arrowheads in B and C.